

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-214265

(43)Date of publication of application : 11.08.1998

(51)Int.Cl.

G06F 17/27

G06F 12/00

(21)Application number : 09-017153

(71)Applicant : MITSUBISHI ELECTRIC CORP

(22)Date of filing : 30.01.1997

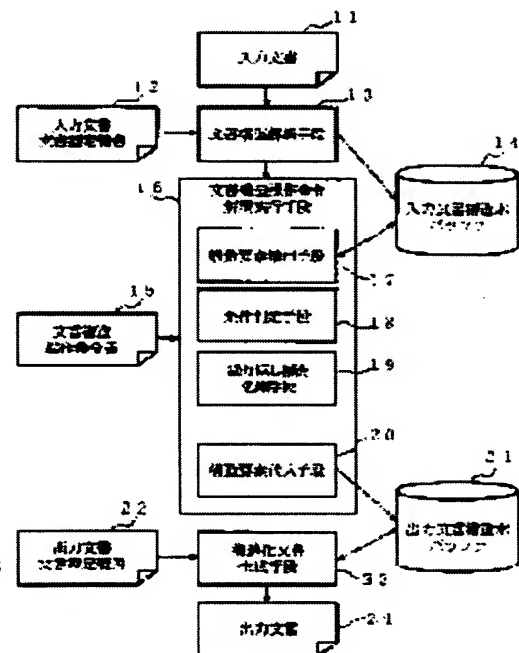
(72)Inventor : IMAMURA MAKOTO
MORIGUCHI OSAMU
SUZUKI KATSUSHI

(54) DEVICE FOR PROCESSING STRUCTURED DOCUMENT

(57)Abstract:

PROBLEM TO BE SOLVED: To realize the processing of a high level structured document such as constraint check between document contents across constituting element identifiers in a document, or conversion, synthesis, and division for plural input and output documents.

SOLUTION: The analysis of the logical structure of an input document 11 is operated, and an input document structural tree is prepared by a document structure analyzing means 13. The reading of a document structure operation instruction 15 is operated, and operation instructions described in it are successively taken out and interpretation-executed by a document structure operation instruction interpretation executing means 16. Then, an output document structural tree is generated from the input document structural tree prepared by the document structure analyzing means 13, and an output document 24 is generated from the prepared output document structural tree by a structured document generating means 23.



LEGAL STATUS

[Date of request for examination] 14.02.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration] abandonment

[Date of final disposal for application] 24.10.2003

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平10-214265

(43)公開日 平成10年(1998) 8月11日

(51)Int.Cl.⁶

G 0 6 F 17/27
12/00

識別記号

5 4 7

F I

G 0 6 F 15/20
12/00

5 5 0 E
5 4 7 H

審査請求 未請求 請求項の数10 O L (全 34 頁)

(21)出願番号 特願平9-17153

(22)出願日 平成9年(1997) 1月30日

(71)出願人 000006013

三菱電機株式会社

東京都千代田区丸の内二丁目2番3号

(72)発明者 今村 誠

東京都千代田区丸の内二丁目2番3号 三

菱電機株式会社内

(72)発明者 森口 修

東京都千代田区丸の内二丁目2番3号 三

菱電機株式会社内

(72)発明者 鈴木 克志

東京都千代田区丸の内二丁目2番3号 三

菱電機株式会社内

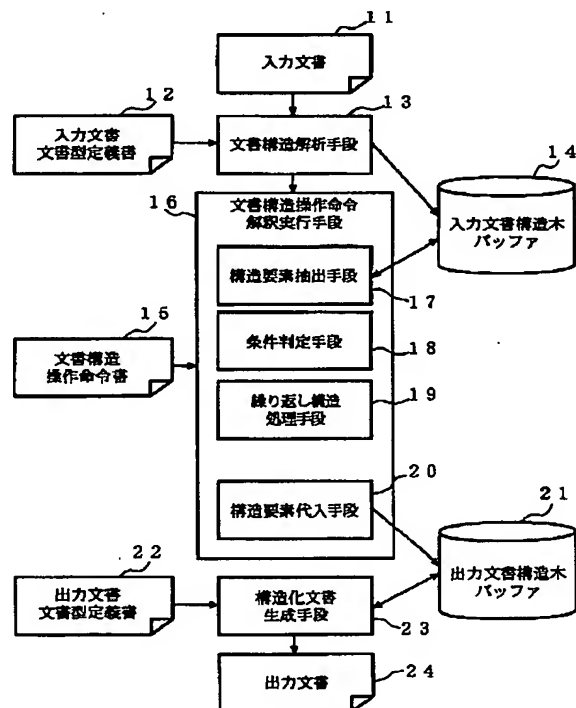
(74)代理人 弁理士 田澤 博昭 (外1名)

(54)【発明の名称】 構造化文書処理装置

(57)【要約】

【課題】 文書内の構成要素識別子にまたがる文書内容間の制約チェックや複数の入出力文書に対する変換・合成・分割といった、高度な構造化文書の処理を実現することができない。

【解決手段】 文書構造解析手段13において入力文書11の論理構造の解析を行なって入力文書構造木を作成し、文書構造操作命令解釈実行手段16において、文書構造操作命令書15の読み込みを行ない、そこに記述されている操作命令を順番に取り出して解釈実行することによって、文書構造解析手段の作成した入力文書構造木より出力文書構造木を生成し、作成された出力文書構造木から構造化文書生成手段23にて出力文書24を生成するようにした。



【特許請求の範囲】

【請求項1】 入力文書の論理構造を解析して、前記入力文書の論理構造を表現する木構造である入力文書構造木を作成する文書構造解析手段と、
前記文書構造解析手段の作成した入力文書構造木を記憶する入力文書構造木バッファと、
文書構造木に対する操作命令が記述された文書構造操作命令書を読み込み、当該文書構造操作命令書に記述された操作命令を順番に取り出して解釈し実行することにより、前記入力文書構造木より出力文書の論理構造を表現する木構造である出力文書構造木を生成する文書構造操作命令解釈実行手段と、
前記文書構造操作命令解釈実行手段の生成した出力文書構造木を記憶する出力文書構造木バッファと、
前記出力文書構造木から出力文書を生成する構造化文書生成手段とを備えた構造化文書処理装置。

【請求項2】 文書構造操作命令解釈実行手段が、読み込んだ文書構造操作命令書に記述されている操作命令を順次取り出し、取り出した前記操作命令のタイプに応じて、
入力文書構造木から構造要素識別子が指定する文書の内容を抽出する構造要素抽出手段、
前記構造要素識別子が指定する文書の内容に課せられる条件により文書の処理内容を制御する条件判定手段、
前記構造要素識別子が繰り返し出現する場合に、その構造要素識別子が指定する各々の文書の内容に対する操作命令を解釈実行する繰り返し構造処理手段、
および前記構造要素識別子が指定する文書内容を出力文書構造木に挿入する構造要素代入手段の中から必要なものを選択して、当該操作命令の解釈実行を行なうものであることを特徴とする請求項1記載の構造化文書処理装置。

【請求項3】 文書構造解析手段から複数の入力文書構造木を受理し、文書構造操作命令書から読み出した入力文書識別子と前記入力文書構造木との対応付けを行なうとともに、構造要素抽出手段から入力文書識別子と構造要素抽出命令を受理すると、その入力文書識別子に対応する入力文書構造木から構造要素識別子が指定する文書の内容を抽出する複数入力文書構造木管理手段を設けたことを特徴とする請求項2記載の構造化文書処理装置。

【請求項4】 構造要素代入手段から出力文書識別子と構造要素代入命令を受理し、前記出力文書識別子に対応する出力文書構造木に、前記構造要素代入命令が指定する文書の内容を挿入する複数出力文書構造木管理手段とを設けたことを特徴とする請求項2または請求項3記載の構造化文書処理装置。

【請求項5】 文書構造操作命令解釈実行手段が、読み込んだ文書構造操作命令書に記述されている操作命令を順次取り出し、取り出した前記操作命令のタイプに応じて、

入力文書構造木から構造要素識別子が指定する文書の内容を抽出する構造要素抽出手段、

前記構造要素識別子が指定する文書の内容に課せられる条件により文書の処理内容を制御する条件判定手段、

前記構造要素識別子が繰り返し出現する場合に、その構造要素識別子が指定する各々の文書の内容に対する操作命令を解釈実行する繰り返し構造処理手段、

前記構造要素識別子が指定する文書の内容を出力文書構造木に挿入する構造要素代入手段、

10 および、辞書ファイルから辞書を読み出し、呼び出された辞書を参照して入力文書に含まれる用語の処理を行なう用語処理手段の中から必要なものを選択して、当該操作命令の解釈実行を行なうものであることを特徴とする請求項1記載の構造化文書処理装置。

【請求項6】 辞書ファイルが、用語検証に必要な用語と、用語に付随する情報を収録した用語辞書を格納した用語辞書ファイルであり、

用語処理手段が、入力文書中の用語が前記用語辞書ファイルから呼び出された用語辞書に収録されている用語か否かを検証し、前記入力文書に含まれる用語に関する誤りの検出や修正を行なう用語検証手段であることを特徴とする請求項5記載の構造化文書処理装置。

【請求項7】 辞書ファイルが、変換前の用語と変換後の用語との対応関係を収録した変換辞書を格納した変換辞書ファイルであり、

用語処理手段が、前記変換辞書ファイルから呼び出された変換辞書を参照して、入力文書に含まれる用語を当該用語に対応付けられている別の用語に変換する用語変換手段であることを特徴とする請求項5記載の構造化文書処理装置。

【請求項8】 定型文書の入力用画面のレイアウトを記述した定型文書入力画面記述書を読み込んで前記定型文書の入力用画面を作成し、定型文書の入力を受理する定型文書入力手段と、

前記定型文書入力画面記述書の記述と出力文書の論理構造との対応関係を記述した定型文書入力画面記述・出力文書構造木対応表を読み込み、前記定型文書入力手段の出力から出力文書構造木を生成する出力文書構造木生成手段とを設けたことを特徴とする請求項1または請求項2記載の構造化文書処理装置。

【請求項9】 文書構造操作命令解釈実行手段が、読み込んだ文書構造操作命令書に記述されている操作命令を順次取り出し、取り出した前記操作命令のタイプに応じて、

入力文書構造木から構造要素識別子が指定する文書の内容を抽出する構造要素抽出手段、

前記構造要素識別子が指定する文書の内容に課せられる条件により文書の処理内容を制御する条件判定手段、

50 前記構造要素識別子が繰り返し出現する場合に、その構造要素識別子が指定する各々の文書の内容に対する操作

命令を解釈実行する繰り返し構造処理手段、
前記構造要素識別子が指定する文書の内容を出力文書構造木に挿入する構造要素代入手段、

および、定型文書入力画面における入力枠と出力文書構造木中の構造要素識別子の対応関係に関する情報を代入することにより、定型文書入力画面記述書の記述と出力文書の論理構造との対応関係を記述した定型文書入力画面記述・出力文書構造木対応表を作成する対応表代入手段の中から必要なものを選択して、当該操作命令の解釈実行を行なうものであることを特徴とする請求項8記載の構造化文書処理装置。

【請求項10】 入力文書の文書型が定義されている入力文書文書型定義書と、出力文書の文書型が定義されている出力文書文書型定義書を読み込んで、それぞれの論理構造を表現する木構造を表示し、両者の前記木構造の構造要素識別子間の対応付けを受理して文書構造操作命令書を作成する入出力文書構造対応付け手段を設けたことを特徴とする請求項1または請求項2記載の構造化文書処理装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、文書作成・承認、既存文書の再利用、文書フォーマットの変更、文書保存・検索・閲覧、データ集計、出版印刷等の文書管理・処理業務において、文書の論理構造を利用して、文書の変換・生成、文書の合成・分割、文書の正しさの検証、さらには文書の誤りの訂正等を行なう構造化文書処理装置に関するものである。

【0002】

【従来の技術】図53は、例えば特開平8-16589号公報に開示される、従来の構造化文書処理装置の構成を示すブロック図であり、構造化文書の持つ誤りを文書構造を考慮しながら校正できるようにしたものである。図において、1はその構造化文書処理装置であり、文書内容が文書の構成要素識別子と対応付けて記述される構造化文書を処理するものである。2は構造化文書処理装置1の備える入力文書ファイルであり、処理対象となる構造化文書を格納するものである。3は同じく構造化文書処理装置1の備える出力文書ファイルであり、処理された構造化文書を格納するものである。

【0003】また、上記構造化文書処理装置1の内部において、4は誤り検出プログラムであり、処理対象となる構造化文書の持つ誤りを検出するものである。5はチェック項目表であり、誤り検出プログラム4の検出処理に必要なチェックデータを管理するものである。6は文書構造表であり、誤り検出プログラム4の抽出した処理対象となる構造化文書の持つ文書構造情報を格納するものである。7はチェック結果表であり、誤り検出プログラム4の検出結果を格納するものである。8は誤り作成プログラムであり、処理対象となる構造化文書の持

つ誤り内容の一覧を作成するものである。9は誤り修正プログラムであり、処理対象となる構造化文書の持つ誤りを修正するものである。

【0004】次に動作について説明する。図53において、誤り検出プログラム4は入力文書ファイル2から読み出された処理対象となる構造化文書を解析すること、当該処理対象となる構造化文書の持つ構成要素識別子と、その構成要素識別子が指定する文書内容とを抽出し、その抽出結果を文書構造表6に格納する。次いで、抽出された文書内容が、チェック項目表5のチェックデータが示す誤り文書条件に該当するか否かをチェックすることで、処理対象となる構造化文書の持つ誤りを検出し、その検出結果をチェック結果表7に格納する。

【0005】図54はそのチェック項目表5の一例を示す説明図である。このチェック項目表5は、処理対象となる構造化文書に対してのチェック条件と、その構造化文書の持つ誤りの修正方法とを構成要素識別子名（図中のチェックIDおよびチェック要素）をキーにして管理するものである。例えば、この図54に示した表中のチェックIDが「1」の行では、「箇条書き」という構成要素識別子が指定する文書内容の要素数が上限値の8個を超えているかどうかをチェックするためのデータを管理している。

【0006】誤り作成プログラム8はこのチェック項目表5に基づいて誤り検出プログラム4が作成したチェック結果表7を参照し、処理対象となる構造化文書の持つ誤り内容の一覧を作成して、それを出力文書ファイル3に出力する。また、誤り修正プログラム9もそのチェック結果表7を参照し、処理対象となる構造化文書の持つ誤りに対して、チェック項目表5のチェックデータが示す修正方法を適用することで誤りの修正を行ない、修正された構造化文書を出力文書ファイル3に出力する。

【0007】なお、このような従来の構造化文書処理装置に関連のある技術についての記載がある文献としては、上記特開平8-16589号公報の他にも、例えば、特開平5-101054号公報、特開平8-137872号公報、特開平8-137877号公報、特開平8-190560号公報などがある。

【0008】

【発明が解決しようとする課題】従来の構造化文書処理装置は以上のように構成されているので、構造化文書に対する操作を記述したチェック項目表5の表現能力が充分ではないため、チェックできる条件に制限があるという課題があった。例えば、構造化文書に対してのチェック条件を構成要素識別子名をキーとして記述しているので、複数の構成要素識別子名にまたがる文書内容間の制約条件をチェックしたり、構成要素識別子の出現の文脈に応じてチェックすべき条件が変わる処理を実行したりするといった処理が困難であったり、チェック項目表5の記述が必要以上に複雑になってしまうなどの問題点が

あった。

【0009】また、チェック項目表5の記述形式が汎用性に欠けているため、誤り修正プログラム9がチェック項目表5の記述形式に依存した形になりやすく、チェックすべき誤りの内容に応じて誤り修正プログラム9の新規作成・変更が必要になったり、複数の入出力文書に対する処理が考慮されておらず、「複数の文書間にまたがる内容の整合性のチェック」や「文書の合成／分割」ができないなどといった課題もあった。

【0010】この発明は上記のような課題を解決するためになされたもので、構造化文書に対する操作命令の集まりである文書構造操作命令書に高い記述能力と汎用性を持たせるとともに、文書構造操作命令書を統一的に解釈実行する文書構造操作命令解釈実行手段を備えることにより、文書内の構成要素識別子にまたがる文書内容間の制約チェックや複数の入出力文書に対する変換・合成・分割といったより高度な構造化文書の処理を実現することのできる構造化文書処理装置を得ることを目的とする。

【0011】また、文書構造操作命令解釈実行手段を文書構造操作命令書中の個々の操作命令に依存しないように実現することにより、「対象とする構造化文書の変更」や「構造化文書に対する操作内容の変更」に対して、文書構造操作命令書のみを変更すればよいようにして、構造化文書処理プログラムの作成の手間を削減可能な構造化文書処理装置を得ることを目的とする。

【0012】

【課題を解決するための手段】請求項1記載の発明に係る構造化文書処理装置は、文書構造解析手段において入力文書の論理構造を解析して入力文書構造木を作成し、文書構造操作命令解釈実行手段において、読み込んだ文書構造操作命令書に記述された操作命令を順番に取り出して解釈実行することで、文書構造解析手段の作成した入力文書構造木より出力文書構造木を生成し、その出力文書構造木から構造化文書生成手段にて出力文書を生成するようにしたものである。

【0013】請求項2記載の発明に係る構造化文書処理装置は、文書構造操作命令解釈実行手段が文書構造操作命令書を読み込んで、構造要素識別子の指定する文書内容を入力文書構造木から抽出する構造要素抽出手段、文書処理内容を構造要素識別子の指定する文書内容に課せられる条件によって制御する条件判定手段、構造要素識別子の指定する各々の文書内容に対する操作命令を解釈実行する繰り返し構造処理手段、および出力文書構造木に構造要素識別子が指定する文書内容を挿入する構造要素代入手段の中から、その読み込んだ文書構造操作命令書より順次取り出した操作命令のタイプに応じて必要なものを選択し、操作命令の解釈実行を行なうようにしたものである。

【0014】請求項3記載の発明に係る構造化文書処理

装置は、複数の入力文書に対する入力文書構造木を管理する複数入力文書構造木管理手段を設けて、文書構造解析手段から複数の入力文書構造木を受理し、入力文書識別子と入力文書構造木を対応付けて入力文書構造木バッファに格納するとともに、構造要素抽出手段から入力文書識別子と構造要素抽出命令を受理した場合に、その入力文書識別子に対応する入力文書構造木から構造要素識別子が指定する文書内容を抽出するようにしたものである。

【0015】請求項4記載の発明に係る構造化文書処理装置は、複数の出力文書に対する出力文書構造木を管理する複数出力文書構造木管理手段を設けて、構造要素代入手段から出力文書識別子と構造要素代入命令を受理した場合に、その出力文書識別子に対応する出力文書構造木に構造要素識別子が指定する文書内容を挿入するようにしたものである。

【0016】請求項5記載の発明に係る構造化文書処理装置は、文書構造操作命令解釈実行手段が文書構造操作命令書を読み込んで、構造要素抽出手段、条件判定手段、繰り返し構造処理手段、構造要素代入手段、さらには、辞書ファイルから辞書呼び出した辞書を参照して入力文書中に含まれる用語の処理を行なう用語処理手段の中から、その読み込んだ文書構造操作命令書より順次取り出した操作命令のタイプに応じて必要なものを選択し、操作命令の解釈実行を行なうようにしたものである。

【0017】請求項6記載の発明に係る構造化文書処理装置は、用語検証に必要な用語と用語に付随する情報を収録した用語辞書を辞書ファイル内に格納しておき、用語処理手段としての用語検証手段により、入力文書中の用語が呼び出された用語辞書に収録されている用語かどうかを検証して、用語に関する誤りの検出や修正を行なうようにしたものである。

【0018】請求項7記載の発明に係る構造化文書処理装置は、変換前の用語と変換後の用語との対応関係を収録した変換辞書を辞書ファイル内に格納しておき、用語処理手段としての用語処理手段により、呼び出された変換辞書を参照することによって、入力文書中に含まれる用語を当該用語に対応付けられている別の用語に変換するようにしたものである。

【0019】請求項8記載の発明に係る構造化文書処理装置は、定型文書の入力用画面のレイアウトを記述した定型文書入力画面記述書を定型文書入力手段で読み込み、定型文書の入力用画面を作成した後、定型文書の入力を受理してその結果を出力文書構造木生成手段に送り、出力文書構造木生成手段では、定型文書入力画面記述書の記述と出力文書の論理構造の対応関係を記述した定型文書入力画面記述・出力文書構造木対応表を読み込んで、定型文書入力手段の出力と定型文書入力画面記述・出力文書構造木対応表の内容から出力文書構造木を生

成し、それを出力文書構造木バッファに格納するようにしたものである。

【0020】請求項9記載の発明に係る構造化文書処理装置は、文書構造操作命令解釈実行手段が文書構造操作命令書を読み込んで、構造要素抽出手段、条件判定手段、繰り返し構造処理手段、構造要素代入手段、さらには、定型文書入力画面における入力枠と出力文書構造木中の構造要素識別子の対応関係に関する情報を定型文書入力画面記述・出力文書構造木対応表に代入して定型文書入力画面記述・出力文書構造木対応表を作成する対応表代入手段の中から、その読み込んだ文書構造操作命令書より順次取り出した操作命令のタイプに応じて必要なものを選択し、操作命令の解釈実行を行なうようにしたものである。

【0021】請求項10記載の発明に係る構造化文書処理装置は、入出力文書構造対応付け手段を設け、入力文書文書型定義書と出力文書文書型定義書とを読み込んで、それぞれの論理構造を表現する木構造を表示し、それら両者の木構造の構造要素識別子間の対応付けを受理して文書構造操作命令書を作成するようにしたものである。

【0022】

【発明の実施の形態】以下、この発明の実施の一形態を説明する。

実施の形態1. ここではまず、与えられた文書構造操作命令書を用いて、入力された構造化文書の文書構造を操作することによって出力文書を生成する、この発明の実施の形態1による構造化文書処理装置の構成と動作について説明する。

【0023】ここで、構造化文書とは、文書の記述内容で区別される表題、著者名、序文、本文のような、文書構成要素の文書中での役割を表現する構造要素識別子を付与することにより、文書の論理的な構造を表現する文書のことである。構造化文書は、文書の構成要素中での役割を表現する構造要素識別子によって文書中の構成要素を確定することができ、またその後の処理のための属性を付与することができるので、文書作成・承認、既存文書の再利用、文書フォーマットの変更、文書保存・検索・閲覧、データ集計、出版印刷等のアプリケーションが必要とする情報を計算機が自動的に抽出することが可能になる。従って、文書を利用した計算機システムより一層の自動化を推進することができる。

【0024】以下、この発明の実施の形態1による構造化文書処理装置を図について説明する。図1はこの実施の形態1の構造化文書処理装置の構成を示すブロック図である。図において、11はこの構造化文書処理装置に入力される入力文書であり、12はこの入力文書の文書型が定義されている入力文書文書型定義書である。ここで、文書型の定義とは、構造化文書において、構造要素識別子によって識別される文書中の構成要素の出現順序

を規定する記述である。13はこれら入力文書11および入力文書文書型定義書12を読み込み、入力文書文書型定義書12に記述されている入力文書11の文書型の定義を参照しながら入力文書11の論理構造を解析した後、当該入力文書11の論理構造を表現する木構造である入力文書構造木を作成する文書構造解析手段である。14はこの文書構造解析手段13が作成した入力文書構造木を記憶する入力文書構造木バッファである。

【0025】15は文書構造木に対する操作命令が記述されている文書構造操作命令書である。この文書構造操作命令書15はこの発明の主要要素であって、文書の変換や文書の正しさの検証など、ユーザが目的とする構造化文書の処理に必要な文書構造に対する手続きを、文書構造の操作命令として記述することが可能になる。16は文書構造操作命令解釈実行手段であり、文書構造操作命令書15を読み込んで文書構造操作命令書15に記述された操作命令を解釈実行し、出力文書の論理構造を表現する木構造である出力文書構造木を入力文書構造木より生成するものである。この文書構造操作命令解釈実行手段16は、上記文書構造操作命令書15と並んでこの発明の主要要素であり、文書構造操作命令書15に記述された操作命令を解釈実行することによって、ユーザが目的とする構造化文書の処理が実現される。

【0026】17は構造要素抽出手段であり、文書構造操作命令書15に記述されている操作命令にしたがって、入力文書構造木バッファ14に格納された入力文書構造木から構造要素識別子が指定する文書の内容を抽出するものである。18は条件判定手段であり、文書構造操作命令書15に記述されている操作命令にしたがって、その構造要素識別子が指定する文書内容に課せられる条件により文書の処理内容を制御するものである。19は繰り返し構造処理手段であり、文書構造操作命令書15に記述されている操作命令にしたがって、文書構造要素識別子が繰り返し出現する場合に、その構造要素識別子が指定するそれぞれの文書内容に対する操作命令を解釈実行するものである。20は構造要素代入手段であり、文書構造操作命令書15に記述されている操作命令にしたがって、構造要素識別子が指定する文書の内容を出力文書構造木に挿入するものである。

【0027】なお、これら構造要素抽出手段17、条件判定手段18、繰り返し構造処理手段19、および構造要素代入手段20は、文書構造操作命令解釈実行手段16によって、文書構造操作命令書15から順次取り出される操作命令のタイプに応じて呼び出されて、当該操作命令の解釈実行に用いられる。

【0028】21は出力文書構造木バッファであり、文書構造操作命令解釈実行手段16による文書構造操作命令書15に記述された操作命令の解釈実行によって作成された出力文書構造木を記憶するものである。22は当該構造化文書処理装置より出力される出力文書の文書型

が定義されている出力文書文書型定義書である。23は構造化文書生成手段であり、出力文書文書型定義書22を読み込み、出力文書文書型定義書22に記述されている出力文書の文書型の定義を参照しながら、出力文書構造木バッファ21に格納されている出力文書構造木より出力文書を生成するものである。24はこの構造化文書生成手段23によって生成、出力される出力文書である。

【0029】図2はこの実施の形態1における入力文書11の一例を示す説明図であり、購入伺い書についてその内容が例示されている。また、図3は同じく出力文書24の一例を示す説明図であり、購入許可書についてその内容が例示されている。なお、これら図2と図3に示された入力文書11および出力文書24はともに構造化文書であり、例えば図2の<伺い書番号>は構造要素識別子の例であり、この購入伺い書の伺い書番号が「314」であることを示している。

【0030】以下、この実施の形態1の説明では、図2に示した購入伺い書に示されている購入希望品の価格の合計が正しいかどうか、また、200000円以上の購入伺い書に必要な部長検印があるかどうかをチェックした後、図3に示した購入許可書を作成して出力する場合を例に説明する。なお、チェックの結果は、図3の購入許可書中の構造要素識別子である<合計チェック>と<検印チェック>の中身に挿入されている。

【0031】図4はこの実施の形態1における入力文書文書型定義書12の内容を示す説明図であり、図2に例示した入力文書11に対する文書型の定義内容の一例が示されている。また、図5は同じく出力文書文書型定義書22の内容を示す説明図であり、図3に例示した出力文書24に対する文書型の定義内容の一例が示されている。なお、これらの文書型定義は、文書中に出現する構造要素識別子の順序を規定するものであり、その意味は、ISO規格8879のSGML (Standard Generalized Markup Language) における文書型定義に準ずるものである。例えば、図4の1行目と2行目は、<購入伺い書>と</購入伺い書>という構造要素識別子に挟まれる領域では、<伺い書番号>、<氏名>、<社員番号>、<購入希望品>、<合計>、<購入希望会社>、そして<部長検印>の順に構造要素識別子が出現することを示している。

【0032】図6はこの実施の形態1における入力文書構造木の一例を示す説明図であり、図2に示した入力文書11に対して文書構造解析手段13が作成したものを例示している。図中の「伺い書番号」等の実線の枠付きの文字列は構造要素識別子であり、構造木は構造要素識別子間の包含関係を上下関係（図中では左右の関係になる）で表現したものである。

【0033】なお、構造木の枝が含む構造要素識別子を図中の左から順に並べたものをパス表現と呼ぶ。例え

ば、図中の点線枠41で囲んで示した枝のパス表現は、<購入伺い書>、<伺い書番号>であり、点線枠42で囲んで示した枝のパス表現は、<購入伺い書>、<購入希望品>、<項目>、<名称>である。また、枝の先にある構造要素を指し示す表現として、「パス表現<購入伺い書>、<伺い書番号>が指し示す構造要素」を用いる。例えば、図6に示す入力文書構造木において、パス表現<購入伺い書>、<伺い書番号>が指し示す構造要素とは「314」であり、パス表現<購入伺い書>、<購入希望品>、<項目>、<名称>が指し示す構造要素とは、「ワープロA」である。

【0034】図7はこの実施の形態1における文書構造操作命令書15の一例を示す説明図であり、文書構造木に対する操作命令の集まりである。操作命令のタイプとしては、構造要素抽出文、条件文、繰り返し文、構造要素代入文などがあり、文書構造操作命令解釈実行手段16では、これらの操作命令のタイプに応じて、構造要素抽出手段17、条件判定手段18、繰り返し構造処理手段19、構造要素代入手段20の中から必要なものを呼び出して、当該操作命令を解釈実行する。なお、図7

(a)には記述された文書構造木に対する操作命令の一例を示しており、同図(b)にはその構造要素抽出文によって抽出される構造要素の一例を示している。

【0035】同図(a)において、点線枠43で囲まれた部分は文書構造操作命令書15の宣言文であり、構造操作の名前が「購入伺い書検印確認」であることを示している。ここでは、構造操作の記述中において、入力ファイルはIで参照され、入力ファイルの文書型定義は「購入伺い書、dtd」であること、および出力ファイルはOで参照され、出力ファイルの文書型定義は「購入許可書、dtd」であることを示している。また、点線枠44で囲まれた部分は変数宣言であり、XとYが変数であること、さらにYが整数であることを示している。

【0036】また、点線枠45で囲まれた部分は構造要素抽出文であり、入力文書構造木中の<購入伺い書>、<伺い書番号>で示されるパス表現が指し示す構造要素を抽出することを示している。図6の説明で述べた例によれば、この場合には「314」を抽出する。点線枠46で囲まれた部分は構造要素代入文であり、前記構造要素抽出文(45)によって抽出された構造要素を、出力文書構造木中のパス表現<購入許可書>、<許可書番号>に対応する枝に挿入することを示す。具体的には、後述する図8に示す出力文書構造木のパス表現<購入許可書>、<許可書番号>に対応する枝に「314」を挿入することを示している。

【0037】点線枠47で囲まれた部分は前記構造要素代入文(46)と同様に、構造要素抽出文を含んだ構造要素代入文であるが、構造要素抽出文中のパス表現<購入伺い書>、<購入希望品>、<項目>、<名称>が入力文書構造木中で繰り返し構造を持つ場合について例示

したものである。この場合には、繰り返し構造を保存して枝を挿入することを意味している。具体的には、図6に示した入力文書構造木中のパス表現<購入伺い書>、<購入希望品>、<項目>、<名称>に対応する枝は、ワープロA、プリンタB、表計算ソフトCの3本があるが、後述する図8に示すように、この構造を保存したまま、出力文書構造木中のパス表現<購入許可書>、<購入許可品>、<品名>に対応する枝を3本作成することを示している。

【0038】点線枠48で囲まれた部分は構造要素抽出文であり、点線枠49で囲まれた部分は繰り返し文である。繰り返し文(49)は構造要素抽出文(48)によって抽出された構造要素の繰り返し構造の中から価格の部分を取り出し、その価格の合計をYに代入することを示している。この場合、構造要素抽出文(48)によって抽出される構造要素は、図7(b)に点線枠52に囲んで示した<名称>と<価格>からなる3つの構造要素が並んだものになり、それが変数Xに代入される。そしてこの構造要素(52)から、繰り返し文(49)中のcar(X)によって、図7(b)の点線枠53、55、56で囲まれたものがその順番で取り出され、car(X)・<価格>によって各々の価格が取り出されて、その合計が変数Yに格納される。

【0039】図7(a)の点線枠50で囲まれた部分は条件文であり、変数Yに格納された購入希望品の合計金額が入力文書構造木中のパス表現<購入伺い書>、<合計>の指し示す文書内容に等しいかどうかをチェックし、等しい場合には出力文書構造木中のパス表現<購入許可書>、<合計チェック>に対応する枝に「成功」を挿入し、そうでない場合には「失敗」を挿入することを示している。また、点線枠51で囲まれた部分も同様の条件文であり、購入伺い書の合計が200000円を超える場合には、購入伺い書の<部長検印>が指し示す文書内容が「あり」かどうかをチェックし、「あり」の場合には出力文書の<検印チェック>に「成功」を挿入し、そうでない場合には<検印チェック>に「失敗」を挿入することを示している。

【0040】このようにして導入したパス表現と条件文とを併用すれば、構成要素識別子の出現の文脈に応じて変わる処理を記述することが可能となる。例えば、図6に例示した入力文書構造木においては、<名称>という構成要素識別子は<購入希望品>および<購入希望会社>に伴って出現しているが、パス表現として<購入伺い書>、<購入希望品>、<項目>、<名称>と、<購入伺い書>、<購入希望会社>、<名称>とを使い分けることによりこの出現の文脈を区別することができる。従って、従来の構造化文書処理装置では困難であった、構成要素識別子の出現の文脈に応じてチェックすべき条件が変わる処理を実行することも可能となる。

【0041】また、同様に、従来の構造化文書処理装置

では困難であった、複数の構成要素識別子名にまたがる文書内容間の制約条件のチェックについても、図7の繰り返し文(49)で示したように、パス表現<購入伺い書>、<購入希望品>、<項目>、<価格>が指し示す数字の合計が、パス表現<購入伺い書>、<合計>が指し示す数字の合計に等しいかどうかというような、複数の構成要素識別子名にまたがる文書内容間の制約条件を記述することができる。

【0042】図8はこの実施の形態1における出力文書構造木の一例を示す説明図であり、図2に例示した入力文書11に対して、文書構造操作命令解釈実行手段16が図7(a)に示した文書構造操作命令書15に記述されている操作命令を解釈実行する過程で、構造要素代入手段20によって作成されたものである。また、図9は同じく出力文書の文書型定義構造木の一例を示す説明図であり、図5に例示した出力文書文書型定義書22に対応する文書型定義構造木を示している。なお、文書型定義構造木の構成方法は入力文書構造木を構成する場合と同様であり、構成要素識別子間の包含関係を上下関係(図中では左右の関係)で表現することにより得られる。

【0043】次に動作について説明する。ここで、図10はこの実施の形態1における構造化文書処理装置の全体の動作を示すフローチャートである。まず最初に、文書構造解析手段13がステップST1において、入力文書11と入力文書文書型定義書12を読み込み、次いでステップST2において入力文書11の文書構造を解析し、解析結果として入力文書構造木を作成して、それを入力文書構造木バッファ14に格納する。例えば、入力文書11が図2に例示したものであり、かつ入力文書文書型定義書12が図4に例示するものであった場合には、図6に例示した入力文書構造木が作成され、その入力文書構造木が入力文書構造木バッファ14に送られて記憶される。

【0044】次に、ステップST3において文書構造操作命令解釈実行手段16は、文書構造操作命令書15の読み込みを行ない、さらにステップST4において、その読み込んだ文書構造操作命令書15からそこに記載されている操作命令を順に取り出す。ここで、文書構造操作命令書15から操作命令が取り出せる場合にはステップST5に進み、取り出せない場合にはステップST10に進む。

【0045】ステップST5では、取り出した操作命令の命令タイプを判定し、その命令タイプに応じて、構造要素抽出手段17、条件判定手段18、繰り返し構造処理手段19、構造要素代入手段20の中の必要なものを呼び出して、取り出された操作命令を解釈実行させる。当該操作命令の解釈実行が終了した後はステップST4に戻り、文書構造操作命令書15から次の操作命令を取り出す。

【0046】ステップST5における判定の結果、命令タイプが構造要素抽出文の場合には構造要素抽出手段17を呼び出し、ステップST6において入力文書構造木から構造要素を抽出する。なお、このステップST6における構造要素抽出の動作の詳細については、図11のフローチャートを用いて後に説明する。

【0047】また、命令タイプが条件文の場合には、条件判定手段18を呼び出してステップST7で条件文を実行し、命令タイプが繰り返し文の場合には、繰り返し構造処理手段19を呼び出してステップST8で繰り返し文を実行する。なお、このステップST7における条件文の実行の処理、およびステップST8における繰り返し文の実行の処理は公知の方法で実現するものとする。

【0048】さらに、命令タイプが構造要素代入文の場合にはステップST9に進み、構造要素代入手段20を呼び出して、出力文書構造木バッファ21に記憶されている出力文書構造木の該当部分に構造要素を挿入する。なお、このステップST9における構造要素挿入の動作の詳細については、図12のフローチャートを用いて後に説明する。

【0049】ここで、例えば入力文書構造木バッファ14中に記憶されている入力文書構造木が図6に例示したものであり、かつ文書構造操作命令書15の内容が図7に例示したものである場合には、文書構造操作命令解釈実行手段16による操作命令の解釈実行の終了時において、図8に例示した出力文書構造木が作成され、それが出力文書構造木バッファ21に格納される。

【0050】ステップST4にて文書構造操作命令書15から操作命令が取り出せなくなると、構造化文書生成手段23はステップST10において、出力文書文書型定義書22の読み込みを行なう。次に構造化文書生成手段23はステップST11において、その読み込んだ出力文書文書型定義22を参照しながら、出力文書構造木バッファ21に格納された出力文書構造木から出力文書24を作成し、ステップST12において、その作成した出力文書24を出力する。なお、ステップST11における出力文書24の作成動作の詳細については、図13のフローチャートを用いて後に説明する。

【0051】図11は上記図10に示したステップST6による構造要素抽出動作の詳細を示すフローチャートである。図10に示すステップST5の判定結果により、文書構造操作命令解釈実行手段16によって呼び出された構造要素抽出手段17が入力文書構造木から構造要素を抽出する場合、まずステップST21において、構造要素抽出文からパス表現の取り出しを行なう。次いでステップST22において、その取り出したパス表現に対応する入力文書構造木中の枝を探した後、ステップST23にて、入力文書構造木中の枝の下にある構造要素を取り出し、処理を図10のステップST4に戻す。

【0052】例えば、入力文書構造木バッファ14中に記憶される入力文書構造木が図6に例示するものであり、かつ構造要素抽出文が図7中に例示された点線枠45で囲まれたものである場合には、パス表現として<購入伺い書>、<伺い書番号>が取り出され、入力文書構造木中の枝として図6の点線枠41で囲まれた部分が選ばれ、構造要素として「314」が取り出される。

【0053】図12は上記図10に示したステップST9による構造要素挿入動作の詳細を示すフローチャートである。図10に示すステップST5の判定結果により、文書構造操作命令解釈実行手段16によって呼び出された構造要素代入手段20が、出力文書構造木の該当部分に構造要素を挿入する場合、まずステップST31において、構造要素代入文の右辺（：=の右側の部分）から代入する構造要素を受け取る。次いでステップST32において、構造要素代入文の左辺（：=の左側の部分）からパス表現を取り出し、ステップST33において、そのパス表現に対応する入力文書構造木中の枝を探した後、ステップST34において、その枝の下にステップST31で得られた構造要素を挿入し、処理を図10のステップST4に戻す。

【0054】例えば、入力文書構造木バッファ14中に記憶されている入力文書構造木が図6に例示したものであり、かつ構造要素代入文が図7(a)に例示された点線枠46で囲まれたものである場合には、代入する構造要素として「314」を受け取り（ステップST31）、パス表現として<購入許可書>、<許可書番号>が取り出され（ステップST32）、出力文書構造木中の枝として図8に例示された点線枠57で囲まれている枝が選ばれ（ステップST33）、構造要素として「314」が挿入される（ステップST34）。

【0055】図13は上記図10に示したステップST11による出力文書24の生成動作の詳細を示すフローチャートである。構造化文書生成手段23が出力文書構造木から出力文書24を作成する場合、まずステップST41において、図10に示すステップST10にて読み込んだ出力文書文書型定義書22より出力文書24の文書型定義構造木を作成する。次いでステップST12において、その文書型定義構造木の文書型定義の枝を左から順に（図9に示した文書型定義構造木では上から順に）たどり、たどる枝がある場合にはステップST43に進み、たどる枝がなくなると図10のステップST12に進む。ステップST43では、文書型定義構造木中の枝に対応する出力文書構造木の枝を探し、その枝に対応する構造要素識別子と構造要素を出力した後、ステップST42に戻り、たどる枝がなくなるまでこの処理を繰り返す。

【0056】例えば、出力文書構造木バッファ21中に記憶されている出力文書構造木が図8に例示されたものであり、かつ出力文書文書型定義書22が図5に例示さ

れたものである場合には、文書型定義構造木として図9に例示されたものがステップST41において作成され、出力文書24として図3に例示した文書が生成される。また、ステップST42において文書型定義構造木の枝を左から順番にたどるとは、図9に例示された出力文書型定義構造木中で、〈購入許可書〉、〈許可書番号〉、〈購入許可書〉、〈検印チェック〉、〈購入許可書〉、〈合計チェック〉、〈購入許可書〉、〈購入許可品〉、〈品名〉、〈購入許可書〉、〈見積り〉、〈社名〉の順番にたどることを意味している。

【0057】このように、図13に示した処理によって、図8に例示した出力文書構造木の枝の順序が、上から、〈購入許可書〉、〈許可書番号〉、〈購入許可書〉、〈購入許可品〉、〈品名〉、〈見積り〉、〈社名〉、〈購入許可書〉、〈合計チェック〉、〈購入許可書〉、〈検印チェック〉の順であるにもかかわらず、図3に例示した出力文書24では、図9に例示した出力文書型定義構造木と同じ順序になっている。

【0058】以上のように、この実施の形態1によれば、図2に示した購入伺い書中の購入希望品の価格の合計金額が当該購入伺い書中の合計と等しいかどうかをチェックするとともに、その合計が200000円以上の場合には部長検印の有無をチェックし、それらのチェック結果を含んだ図3に示す購入許可書を出力文書として生成することができ、このように、高い表現能力を持つ文書構造操作命令書15に記述された文書構造に対する操作命令を解釈実行することにより、複数の構成要素識別子名にまたがる文書内容間の制約条件のチェックや、構成要素識別子の出現の文脈に応じてチェックすべき条件が変わる処理の実行など、より高度な構造化文書処理を実現することが可能となり、さらに、文書構造操作命令解釈実行手段16を文書構造操作命令書15中の個々の操作命令に依存しないように実現することができるので、対象とする構造化文書の変更や、構造文書の操作内容の変更に対しても、文書構造操作命令書15のみを変更するだけでよくなるため、構造化文書処理プログラムの作成の手間を削減することが可能になるなどの効果がある。

【0059】実施の形態2。上記実施の形態1では、1つの入力文書について出力文書を生成するものについて説明したが、複数の入力文書に対して出力文書を生成することも可能である。この実施の形態2はそのような複数の入力文書に対して出力文書を生成する構造化文書処理装置に関するものであり、これにより、一方の入力文書に記載されている内容と、もう一方の入力文書に記載されている内容とに矛盾がないかどうかをチェックしたり、複数の入力文書を合成して一つの出力文書を作成することができるようになる。

【0060】以下、そのようなこの発明の実施の形態2を図について説明する。図14はこの実施の形態2によ

る構造化文書処理装置の構成を示すブロック図であり、実施の形態1と同一または相当する部分については、図1と同一符号を付しその説明を省略する。図において、25は複数入力文書構造木管理手段であり、文書構造解析手段13から複数の入力文書構造木を順番に受理し、文書構造操作命令書15から読み出した入力文書識別子と入力文書構造木とを対応付けて入力文書構造木バッファ14に格納するものである。また、構造要素抽出手段17から入力文書識別子と構造要素抽出命令を受理した場合10には、その入力文書識別子に対応する入力文書構造木から構造要素識別子が指定する文書の内容を抽出する。

【0061】図15はこの実施の形態2における入力文書11の一例を示す説明図であり、この場合には従業員名簿の記載例が示されている。この従業員名簿においては従業員の氏名と社員番号が記載されており、例えば、氏名「森口 太郎」の社員番号は「15926」であり、氏名「今村 次郎」の社員番号は「53589」であることを示している。

【0062】図16はこの実施の形態2における、入力文書識別子と入力文書構造木の対応関係を表現する入力文書識別子・入力文書構造木対応表の一例を示す説明図である。この入力文書識別子・入力文書構造木対応表は、複数入力文書構造木管理手段25が、文書構造操作命令書15から読み出した入力文書識別子と入力文書構造木とを対応付けて入力文書構造木バッファ14に格納する際のデータ構造を示したものである。なお、図中の点線枠61で囲まれた部分は図2に例示した入力文書11に対する入力文書構造木、点線枠62で囲まれた部分は図15に例示した入力文書11に対する入力文書構造木であり、点線枠63で囲まれた部分は上記入力文書構造木(61)中の枝、点線枠64で囲まれた部分は上記入力文書構造木(62)中の枝である。

【0063】図17はこの実施の形態2における文書構造操作命令書15の一例を示す説明図であり、図2に例示した購入伺い書、および図15に例示した従業員名簿を入力文書11として受理して、購入伺い書における氏名と社員番号の対応関係が、従業員名簿に記載される氏名と社員番号の対応関係と同じであるかどうかをチェックするために必要な文書構造の操作命令が記載されている場合について例示している。

【0064】なお、図中の点線枠65で囲まれた部分は入力文書識別子とその入力文書識別子に対応する文書型定義の対応関係を記述しており、入力文書識別子I1に対応する入力文書の文書型定義の名称が「購入伺い書、dtd」であること、および入力文書識別子I2に対応する入力文書の文書型定義の名称が「従業員名簿、dtd」であることを示している。また、点線枠66で囲まれた部分は入力文書識別子I1に対応する入力文書構造木に対する構造要素抽出文であり、図16に示した入力

文書構造木(61)中の点線枠63で囲まれた枝の下
の構造要素を抽出することを示している。同様に、点線枠
67で囲まれた部分は入力文書識別子12に対応する入
力文書構造木に対する構造要素抽出文であり、図16に
示す入力文書構造木(62)中の点線枠64で囲まれた
枝の下の構造要素を抽出することを示している。

【0065】このように、文書構造操作命令書15にお
ける構造要素抽出文に入力文書識別子を埋め込むこと
により、入力文書識別子・入力文書構造木対応表中のど
の入力文書構造木から構造要素を抽出するかを表現する
ことができ、従来の構造化文書処理装置では困難であ
った、複数の文書間にまたがる内容の整合性のチェック
や文書の合成を行なうことも可能となる。

【0066】次に動作について説明する。ここで、図1
8は文書構造操作命令解釈実行手段16によって呼び出
された構造要素抽出手段17により、入力文書構造木パ
ッファ14に記憶されている入力文書識別子・入力文書
構造木対応表中の入力文書構造木から構造要素を抽出
する際の動作を示すフローチャートである。図10のフ
ローチャートに示したステップST5における判定の結果
により、構造要素抽出手段17が、入力文書識別子・入
力文書構造木対応表中の入力文書構造木から構造要素
を抽出する場合、まずステップST51において、構造要
素抽出文から入力文書識別子とパス表現の取り出しを行
なう。次いでステップST52において、その取り出し
た入力文書識別子に対応する入力文書構造木に対して、
取り出したパス表現に対応する枝を探す。その後ステッ
プST53において、入力文書構造木中の上記ステップ
ST52で得られた枝の下にある構造要素を取り出し
て、処理を図10のステップST4に戻す。

【0067】例えば、入力文書構造木バッファ14中に
記憶されている入力文書識別子・入力文書構造木対応表
が図16に例示するものであり、かつ構造要素抽出文が
図17に例示された点線枠65で囲まれたものである場
合には、ステップST51においてパス表現として<購
入伺い書>、<社員番号>が取り出され、ステップST
52において入力文書構造木中の枝として、図16中の
点線枠63で囲まれた部分選ばれ、ステップST53
において構造要素「15926」が取り出される。

【0068】以上のように、この実施の形態2によ
れば、図2に示した購入伺い書と図15に示した従業員名
簿とを入力文書11として受理し、購入伺い書における
氏名と社員番号の対応関係が、従業員名簿に記載され
る氏名と社員番号の対応関係と同じであるかどうかを
チェックすることができ、このように、入力文書識別子・入
力文書構造木対応表により構造要素を抽出すべき入力文
書構造木を選択することが可能になるので、一方の文書
に記載されている内容ともう一方の文書に記載されて
いる内容とに矛盾があるか否かをチェックしたり、複数
の文書を合成して一つの文書を作成するといったより高度

な構造化文書処理を実現することが可能になるので、構
造化文書に対する文書変換や文書の正しさの検証等の能
力を向上させることができる効果がある。

【0069】実施の形態3。上記実施の形態2では、複
数の入力文書に対して出力文書を生成するものについて
説明したが、入力文書から複数の出力文書を生成するこ
とも可能である。この実施の形態3はそのような入力文
書から複数の出力文書を生成する構造化文書処理装置に
関するものであり、これによって、文書中の構造要素が
満たすべき条件に応じて文書を分割することができるよ
うになり、また、上記実施の形態2の構造化文書処理装
置と組み合わせることにより、複数の入力文書から複数
の出力文書を生成することができるようになる。

【0070】以下、そのようなこの発明の実施の形態3
を図について説明する。図19はこの実施の形態3によ
る構造化文書処理装置の構成を示すブロック図であり、
実施の形態1と同一または相当する部分については、図
1と同一符号を付しその説明を省略する。図において、
26は複数出力文書構造木管理手段であり、構造要素代
入手段20から出力文書識別子と構造要素代入命令を受
理した場合に、その出力文書識別子に対応する出力文書
構造木にその構造要素代入命令が指定する文書の内容を
挿入するものである。

【0071】図20は図3に例示した購入許可書を入力
文書11とする場合の出力文書24の一例を示す説明図
である。なお、同図(a)は図3に示す購入許可書か
ら、購入品の価格が200000円以上の購入品の名称
と価格のリストを抽出、記載した出力文書24について
例示したものである。同様に、図20(b)は図3に示
す購入許可書から、購入品の価格が200000円未満
の購入品の名称と価格のリストを抽出、記載した出力文
書24について例示したものである。

【0072】図21はこの実施の形態3における、出力
文書識別子と出力文書構造木の対応関係を表現する入力
文書識別子・入力文書構造木対応表の一例を示す説明図
である。この出力文書識別子・出力文書構造木対応表
は、構造要素代入手段20が文書構造操作命令書15に
記載された構造要素代入文にしたがって出力文書構造木
を作成する際に用いられるものであり、出力文書構造木
バッファ21に記憶されているデータ構造の一例につい
て例示したものである。なお、図中の点線枠71で囲ま
れた部分は図20(a)に示した出力文書24を生成す
るための元となる出力文書構造木であり、点線枠72で
囲まれた部分は図20(b)に示した出力文書24を生成
するための元となる出力文書構造木である。従って、
構造化文書生成手段23により、この出力文書構造木
(71)からは図20(a)に示す出力文書24が生成
され、出力文書構造木(72)からは図20(b)に示
す出力文書24が生成される。また、点線枠73で囲ま
れた部分は上記出力文書構造木(71)中の枝、点線枠

74で囲まれた部分は上記出力文書構造木(72)中の枝である。

【0073】図22はこの実施の形態3における文書構造操作命令書15の一例を示す説明図であり、図3に例示した購入許可書を入力文書11として受理し、当該購入許可書中の購入品の価格が200000円以上であるものとそうでないものに選別した後、各々の条件を満たす購入品のリストを別文書として出力するために必要な文書構造の操作命令が記載されている場合について例示している。より具体的には、購入品の価格が200000円以上の場合には、出力文書識別子01に対応する出力文書構造木に購入品の品名と価格を挿入し、200000円未満の場合には、出力文書識別子02に対応する出力文書構造木に購入品の品名と価格を挿入する操作命令が記載されている。

【0074】なお、図中の点線枠75で囲まれた部分は出力文書識別子に対応する文書型定義の対応関係を記述しており、出力文書識別子01と02に対応する入力文書の文書型定義の名称が「購入品リスト、dtd」であることを示している。また、点線枠76で囲まれた部分は出力文書識別子01に対応する出力文書構造木に対する構造要素代入文であり、図21に示した出力文書構造木(71)中の点線枠73で囲まれた枝の下に「200000以上」という構造要素を挿入することを示している。同様に、点線枠77で囲まれた部分は出力文書識別子02に対応する出力文書構造木に対する構造要素代入文であり、図21に示した出力文書構造木(72)中の点線枠74で囲まれた枝の下に「200000未満」という構造要素を挿入することを示している。

【0075】このように、文書構造操作命令書15における構造要素代入文に出力文書識別子を埋め込むことにより、出力文書識別子・出力文書構造木対応表中のどの出力文書構造木に対して構造要素を挿入するかを表現することができ、従来の構造化文書処理装置では困難であった、複数の文書間にまたがる内容の整合性のチェックや文書の分割を行なうことも可能となる。

【0076】次に動作について説明する。ここで、図23は文書構造操作命令解釈実行手段16によって呼び出された構造要素代入手段20により、出力文書構造木バッファ21に記憶されている出力文書識別子・出力文書構造木対応表中の出力文書構造木に構造要素を挿入する際の動作を示すフローチャートである。図10のフローチャートに示したステップST5における判定の結果により、構造要素代入手段20が出力文書識別子・出力文書構造木対応表中の出力文書構造木に構造要素を挿入する場合、まずステップST61において、構造要素代入文の右辺(：＝の右側の部分)から代入する構造要素を受け取り、さらにステップST62において、左辺(：＝の左側の部分)から出力文書識別子とパス表現を取り出す。次いでステップST63において、その取り出し

た出力文書識別子に対応する出力文書構造木に対して、取り出したパス表現に対応する枝を探す。その後ステップST64において、上記ステップST63で得られた枝の下に、上記ステップST61で得られた構造要素を挿入して、処理を図10のステップST4に戻す。

【0077】例えば、出力文書構造木バッファ21に記憶される出力文書識別子・出力文書構造木対応表が図21に例示されたものであり、かつ構造要素代入文が図22に例示された点線枠76で囲まれたものである場合には、ステップST61において構造要素として「200000以上」を受け取り、ステップST62においてパス表現として<購入品リスト>、<タイプ>が取り出される。そして、ステップST63において出力文書構造木の枝として図21中の点線枠73で囲まれた部分を選ばれ、ステップST64において構造要素「200000以上」が代入される。

【0078】以上のように、この実施の形態3によれば、図3に示した購入許可書を入力文書11として受理し、購入許可書中の購入品の価格が200000円以上である購入品のリストである、図20(a)に示した出力文書24と、購入許可書中の購入品の価格が200000円未満である購入品のリストである、図20(b)に示した出力文書を別々に生成することができ、このように、出力文書識別子・出力文書構造木対応表により構造要素を挿入すべき出力文書構造木を選択することが可能になるので、文書中の構造要素が満たすべき条件に応じて文書を分割したり、複数の入力文書11から複数の出力文書24を生成するといったより高度な構造化文書処理を実現することが可能となるので、構造化文書に対する文書変換や文書の正しさの検証等の能力を向上させることができる効果がある。

【0079】実施の形態4. 上記各実施の形態においては、操作命令のタイプが構造要素抽出文、条件文、繰返し文、および構造要素代入文であった場合について説明したが、操作命令のタイプにさらに辞書呼び出し文を追加することも可能である。実施の形態4はその辞書呼び出し文を用いて入力文書中に含まれる用語に関する誤りの検出を行ない、その検出結果を含む文書を生成する構造化文書処理装置に関するものである。

【0080】以下、そのようなこの発明の実施の形態4を図について説明する。図24はこの実施の形態4による構造化文書処理装置の構成を示すブロック図であり、実施の形態1と同一または相当する部分については、図1と同一符号を付しその説明を省略する。図において、27は辞書が格納された辞書ファイルとしての用語辞書ファイルであり、その辞書として用語検証に必要な用語と用語に付随する情報を収録した用語辞書が格納されたものである。28は用語処理手段としての用語検証手段であり、文書構造操作命令解釈実行手段16から呼び出され、文書構造操作命令書15中の操作命令にしたがっ

て、入力文書11中の用語が用語辞書ファイル27から呼び出した用語辞書に収録されている用語かどうかを検証するものである。29は用語辞書管理手段であり、用語検証手段28から用語辞書の検索命令を受理すると、用語辞書ファイル27から呼び出された用語辞書を検索して、その検索結果を用語検証手段28に返すものである。なお、この用語辞書管理手段29は公知の方法で実現するものとし、詳細な構成と動作の記述はここでは省略する。

【0081】図25はこの実施の形態4における入力文書11の一例を示す説明図、図26は同じく出力文書24の一例を示す説明図である。なお、図25に示す入力文書11は参考文献のリストを記載した文書を例示したものであり、参考文献ごとにタイトルと雑誌名称を記載している。また、図26に示す出力文書24は図25に示した参考文献のリストにおける雑誌名称の誤りを記載した文書を例示したものであり、参考文献中の雑誌名称の誤りとその誤りの内容を示すエラーメッセージを記載している。

【0082】図27はこの実施の形態4における用語辞書ファイル27より呼び出される用語辞書の内容の一例を示す説明図である。この用語辞書はその名称(辞書名)を「雑誌名称辞書」とする、雑誌の名称を収録したものについて例示したものであり、「情報処理学会全国大会」や「情報処理学会論文誌」などはその雑誌の名称の一例を示したものである。

【0083】図28はこの実施の形態4における文書構造操作命令書15の一例を示す説明図であり、図25に例示した参考文献のリストを入力文書11として受理し、その参考文献の雑誌名称が図27に例示した「雑誌名称辞書」に収録されているものかどうかをチェックし、収録されていない場合にはその雑誌名称とエラーメッセージを含んだ出力文書24を生成するために必要な文書構造の操作命令が記載されている場合について例示している。

【0084】図において、点線枠81で囲まれた部分は辞書識別子とその辞書識別子に対応する辞書名との対応関係を記述しており、辞書識別子D1の辞書名が「雑誌名称辞書」であることを示している。点線枠82および点線枠83で囲まれた部分は構造要素代入文であり、変数Yにはこれら構造要素代入文(82)および構造要素代入文(83)によって、図25に例示した入力文書11中でパス表現<参考文献>、<項目>が指し示す構造要素が順番に代入されることを示している。点線枠84で囲まれた部分は辞書呼び出し文であり、文の左辺(属することを示す数学記号の左側)のY、<雑誌名称>が指し示す用語が、文の右辺(属することを示す数学記号の右側)の辞書識別子D1が示す雑誌名称辞書に収録されているかどうかをチェックする。変数Yには前述のように、図25の入力文書11中でパス表現<参考文献

>、<項目>が指し示す構造要素が順番に代入されるので、結局、当該辞書呼び出し文(84)の左辺であるY、<雑誌名称>には、入力文書11中の参考文献の雑誌名称が順番に代入されることになり、この辞書呼び出し文(84)によって、参考文献の雑誌名称が図27に例示した雑誌名称辞書に収録されているかどうかをチェックする文書構造の操作命令を表現することができる。

【0085】次に動作について説明する。ここで、図29はこの実施の形態4における構造化文書処理装置の全体の動作を示すフローチャートである。なお、基本的な動作は実施の形態1において説明したものと同等であるためここではその説明は省略し、それとは異なる部分を中心に説明する。ステップST5において、取り出した操作命令の命令タイプを判定し、その命令タイプに応じて、構造要素抽出手段17、条件判定手段18、繰り返し構造処理手段19、構造要素代入手段20、用語検証手段28の中の必要なものを呼び出して、取り出された操作命令を解釈実行させる。当該操作命令の解釈実行が終了した後はステップST4に戻り、文書構造操作命令書15から次の操作命令を取り出す。

【0086】ステップST5における判定の結果、命令タイプが構造要素抽出文であれば構造要素抽出手段17を呼び出してステップST6で入力文書構造木から構造要素を抽出し、条件文であれば条件判定手段18を呼び出してステップST7で条件文を実行し、命令タイプが繰り返し文であれば繰り返し構造処理手段19を呼び出してステップST8で繰り返し文を実行し、構造要素代入文であれば構造要素代入手段20を呼び出してステップST9で出力文書構造木に構造要素を挿入する。また、ステップST5における判定の結果、命令タイプが辞書呼び出し文であった場合には用語検証手段28を呼び出して、ステップST13において入力文書11中の用語が用語辞書ファイル27より呼び出した用語辞書に収録されている用語かどうかの検証を行なう。

【0087】図30はこのステップST13における文書構造操作命令書15中の辞書呼び出し文を解釈実行する動作を示すフローチャートである。文書構造操作命令解釈実行手段16に呼び出された用語検証手段28は、まずステップST71において、用語辞書ファイル27より辞書呼び出し文の右辺にある用語辞書の呼び出しを行なう。次いでステップST72において、当該辞書呼び出し文の左辺にある用語が、ステップST71で用語辞書ファイル27より呼び出した用語辞書に収録されているか否かのチェックを行なう。その結果、当該用語がその用語辞書に収録されている場合には、ステップST73に進んで真(true)を返した後、処理を図29のステップST4に戻す。一方、収録されていない場合にはステップST74に進んで偽(false)を返した後、上記ステップST4に処理を戻す。

【0088】例えば、入力文書11が図25に例示した

ものであり、かつ辞書呼び出し文が図28中に例示された点線枠84で囲まれたものであり、かつ用語辞書が図27に例示したものである場合には、ステップST71で図27に示す雑誌名称辞書が用語辞書ファイル27から呼び出され、ステップST72で図25に示す入力文書中の参考文献の雑誌名称である「情報処理学会全国大会」と「情処論」が当該雑誌名称辞書に収録されているかどうかのチェックをする。「情報処理学会全国大会」は図27に示す雑誌名称辞書に収録されているのでステップST73で真(true)を返し、「情処論」は図27に示す雑誌名称辞書には収録されていないのでステップST74で偽(false)を返す。このステップST74で偽(false)が返された場合の出力文書24の例が図26に示されている。

【0089】以上のように、この実施の形態4によれば、図25に例示した参考文献リストを入力文書11として受理し、参考文献の雑誌名称が図27に示した雑誌名称辞書に収録されているものかどうかをチェックして、収録されていない場合にはその雑誌名称とエラーメッセージを含んだ、図26に例示する出力文書24を生成することができ、このように、辞書呼び出し文を解釈実行する用語検証手段28により、入力文書11中に含まれる用語が用語辞書に含まれるかどうかをチェックすることが可能となるため、入力文書11中に含まれる用語に関する誤りを検出して、その誤り検出の結果を含む出力文書24を生成するといったより高度な構造化文書処理を実現することができる効果がある。

【0090】実施の形態5。上記実施の形態4においては、操作命令のタイプに辞書呼び出し文を追加し、それを用いて入力文書中に含まれる用語に関する誤りの検出を行なう場合について説明したが、その辞書呼び出し文を用いて、入力文書に含まれる用語を別の用語に変換した後、出力文書を生成することも可能である。図31はそのようなこの発明の実施の形態5による構造化文書処理装置の構成を示すブロック図であり、実施の形態1と同一または相当する部分については、図1と同一符号を付しその説明を省略する。

【0091】図において、30は辞書が格納された辞書ファイルとしての変換辞書ファイルであり、辞書として変換前の用語と変換後の用語の対応関係を収録した変換辞書が格納されているものである。31は用語処理手段としての用語変換手段であり、文書構造操作命令解釈実行手段16から呼び出され、文書構造操作命令書15中の操作命令にしたがって、入力文書11中の用語をそれに対応する別の用語に変換するものである。32は用語辞書管理手段であり、変換対象となる用語を受理すると、変換辞書ファイル30から呼び出された変換辞書を検索して、その用語に対応する別の用語に変換し、その変換によって得られた用語を用語変換手段31に返すものである。なお、この変換辞書管理手段32は公知の方

法で実現するものとし、その詳細な構成と動作の記述はここでは省略する。

【0092】図32はこの実施の形態5における入力文書11の一例を示す説明図、図33は同じく出力文書24の一例を示す説明図である。なお、図32に示す入力文書11は参考文献のリストを記載した文書を例示したものであり、参考文献ごとにタイトルと雑誌名称を記載している。また、図33に示す出力文書24は、用語変換手段31が変換辞書ファイル30より呼び出した変換辞書を参照することにより、図32に例示した入力文書11の雑誌名称を変換した後、出力した文書を例示したもので、図32に示した入力文書11における参考文献の雑誌名称「情処論」が「情報処理学会論文誌」に変換されている。

【0093】図34はこの実施の形態5における変換辞書ファイル30より呼び出される変換辞書の内容の一例を示す説明図である。この変換辞書はその名称(辞書名)を「雑誌略称・正式名称対応辞書」とし、変換前の用語として雑誌の略称を収録し、変換後の用語として雑誌の正式名称を収録している。例えば、「情処全大」の正式名称は「情報処理学会全国大会」であり、「情処論」の正式名称は「情報処理学会論文誌」であることを示している。

【0094】図35はこの実施の形態5における文書構造操作命令書15の一例を示す説明図であり、図32に例示した参考文献のリストを入力文書11として受理し、その参考文献の雑誌名称が図34に例示した変換辞書「雑誌略称・正式名称対応辞書」の変換前の用語(略称)として収録されているかどうかをチェックし、収録されている場合にはその雑誌名称を変換後の用語(正式名称)に変換して出力文書24を生成するために必要な文書構造の操作命令が記載されている場合について例示している。

【0095】図において、点線枠85で囲まれた部分は図28の点線枠81で囲まれた部分と同様に、辞書識別子とその辞書識別子に対応する辞書名との対応関係を記述しており、辞書識別子D1の辞書名が「雑誌名称略称・正式名称対応辞書」であることを示している。点線枠86で囲まれた部分は変換辞書を呼び出す辞書呼び出し文であり、文の左辺(属することを示す数学記号の左側)Y、＜雑誌名称＞が指し示す用語が、文の右辺(属することを示す数学記号の右側)の辞書識別子D2が指し示す「雑誌名称・略称正式名称対応辞書」に変換前の用語として収録されているかどうかをチェックし、収録されている場合にはその変換前の用語に対応付けられた変換後の用語を変数Zに代入することを示している。点線枠87で囲まれた部分は構造要素代入文であり、辞書呼び出し文(86)によって得られた変換後の用語Zを出力文書識別子Oが指し示す出力文書構造木のパス表現＜参考文献＞、＜項目＞、＜雑誌名＞が示す枝の下に挿

入することを示している。従って、この辞書呼び出し文(86)によって、参考文献の雑誌名称が略称である場合には、それを正式名称に変換する文書構造の操作命令を表現することができる。

【0096】次に動作について説明する。ここで、図36はこの実施の形態5における構造化文書処理装置の、文書構造操作命令書15中で変換辞書を読み出す辞書呼び出し文を解釈実行する動作を示すフローチャートである。なお、全体動作については、図29にそのフローチャートを示した実施の形態4の場合と同様であるため、

ここではその説明は省略する。

【0097】まずステップST81において、変換辞書ファイル30より辞書呼び出し文の右辺にある変換辞書の呼び出しを行なう。次いでステップST82において、当該辞書呼び出し文の左辺にある用語が、ステップST81で変換辞書ファイル30より呼び出した変換辞書に収録されているか否かのチェックを行なう。その結果、当該用語がその変換辞書に収録されている場合にはステップST83に進んで、辞書呼び出し文の右辺にある変数に変換後の用語を代入した後、真(true)を返して処理を図29のステップST4に戻す。一方、収録されていない場合にはステップST84に進んで偽(false)を返した後、上記ステップST4に処理を戻す。

【0098】例えば、入力文書11が図32に例示したものであり、かつ辞書呼び出し文が図35中に例示された点線枠86で囲まれたものであり、かつ変換辞書が図34に例示したものである場合には、ステップST81で図34に示した「雑誌名称・正式名称対応辞書」が変換辞書ファイル30から呼び出され、ステップST82にて図32に示す入力文書11中の参考文献の雑誌名称である「情報処理学会全国大会」と「情処論」が当該「雑誌名称・正式名称対応辞書」に収録されているかどうかのチェックが行なわれる。その結果、「情報処理学会全国大会」は図34に示した「雑誌名称・正式名称対応辞書」に変換前の用語として収録されていないので、ステップST83において偽(false)を返し、「情処論」は図34に示した「雑誌名称・正式名称対応辞書」に変換前の用語として収録されているので、ステップST84において、図35の辞書呼び出し文(86)の変数Zに当該「情処論」に対応付けられた変換後の用語「情報処理学会論文誌」を代入した後、真(true)を返す。

【0099】以上のように、この実施の形態5によれば、図32に例示した参考文献リストを入力文書11として受理して、参考文献の雑誌名称が図34に示した変換辞書(雑誌略称・正式名称対応辞書)に変換前の用語(略称)として収録されているものかどうかをチェックし、収録されている場合にはその雑誌名称を変換後の用語(正式名称)に変換して出力文書を生成することがで

き、このように、変換辞書を読み出す辞書呼び出し文を解釈実行する用語変換手段31により、入力文書11中に含まれる用語を対応する別の用語に変換することが可能となるため、入力文書11中に含まれる用語に関する誤りを修正した文書を生成したり、用語の表記を変更した文書を生成するといった、より高度な構造化文書処理を実現することができる効果がある。

【0100】実施の形態6. この発明による構造化文書処理装置は、定型文書の入力画面を利用して構造化文書を編集するようにすることも可能である。図37はそのようなこの発明の実施の形態6による構造化文書処理装置の構成を示すブロック図であり、実施の形態1と同様または相当する部分については、図1と同一の符号を付しその説明を省略する。

【0101】図において、33は定型文書入力画面記述書であり、定型文書の入力用画面のレイアウトを記述したものである。34は定型文書入力手段であり、定型文書入力画面記述書33を読み込み、定型文書の入力用画面を作成した後、ユーザからの定型文書入力を受理して、その結果を後述する出力文書構造木生成手段36に送るものである。なお、この定型文書画面記述書33と定型文書入力手段34とは公知の方法で実現するものとし、ここではその詳細構成と動作の記述は省略する。35は定型文書入力画面記述・出力文書構造対応表であり、定型文書入力画面記述書33に記述された定型文書の入力用画面のレイアウトと出力文書24の論理構造との対応関係を記述したものである。36は出力文書構造木生成手段であり、定型文書入力画面記述・出力文書構造対応表35を読み込んで、定型文書入力手段34の出力から出力文書構造木を生成し、それを出力文書構造木バッファ21に格納するものである。

【0102】図38はこの実施の形態6における定型文書入力画面記述書33の内容の一例を示す説明図であり、氏名を「森口 太郎」、「今村 次郎」、「藤井 三郎」の中から選択するメニューと、「品名」と「価格」という枠に文字列の入力を促す画面を表現したものを例示している。なお、その記述形式と意味は、例えば公知のWWW(World Wide Web)の文書記述言語HTML(Hyper Text Markup Language)に準ずるものとする。

【0103】図39は上記図38に例示した定型文書入力画面記述書33から定型文書入力手段34によって作成される、この実施の形態6における定型文書入力画面の一例を示す説明図である。なお、この定型文書入力手段34による定型文書入力画面の生成動作は、例えば公知の技術であるWWWのブラウザの動作に準ずるものとする。

【0104】図40は図39に例示した定型文書入力画面に対してユーザがメニュー選択と文字列入力をした結果、定型文書入力手段34より送出される出力の一例を

示す説明図である。この定型文書入力手段34の出力は同図に点線枠91で囲んで示すような式の集まりである。なお、図中の&は式の始まりを示す記号であり、この発明では、記号=の左辺を属性と呼び、右辺をその属性の値と呼ぶ。点線枠91で囲まれた式の例によれば、「1. 2」が属性であり、属性「1. 2」の値は「森口 太郎」である。なお、この定型文書入力手段34の出力中の式の属性は、定型文書入力画面におけるメニューを選択する枠や文字列を入力する枠（以後入力枠と呼ぶ）を同定する識別子の役割を持つ。

【0105】図41はこの実施の形態6における定型文書入力画面記述・出力文書構造記述対応表35の一例を示す説明図であり、定型文書入力画面入力枠の識別子を構成する数字（以下、入力枠識別子と呼ぶ）と出力文書構造木における構造要素識別子との対応関係を記述している。この定型文書入力画面記述・出力文書構造記述対応表35を用いることにより、定型文書の入力画面中の入力枠と出力文書構造木中の枝に対応するパス表現とを対応付けることができる。例えば、図40の点線枠91で囲まれた式の属性「1. 2」に対応するパス表現は、

入力枠識別子の1に対応する構造要素識別子が、図41に点線枠92で囲んで示すように<購入伺い書>であり、入力枠識別子の2に対応する構造要素識別子が、点線枠93で囲んで示すように<氏名>であることから、<購入伺い書>、<氏名>であることがわかる。

【0106】図42は出力文書構造木生成手段36によって生成される、この実施の形態6における出力文書構造木の一例を示す説明図である。図40の点線枠91で囲まれた式の属性「1. 2」に対応する枝は、図42に点線枠94で囲んで示したものとなる。なぜなら、図41の説明で述べたように、属性「1. 2」に対応するパス表現は<購入伺い書>、<氏名>となるからである。従って、定型文書入力画面記述・出力文書構造対応表35を用いることにより、定型文書入力手段34の出力中の式の属性と出力文書構造木の枝とを対応付けることが可能となる。また、実施の形態1において説明した動作と同様にして、構造化文書生成手段23によってこの文書構造木から図2に例示したものと同様の出力文書24を生成することができる。

【0107】次に動作について説明する。ここで、図43は上記出力文書構造木生成手段36の動作を示すフローチャートである。まずステップST91において、定型文書入力手段34の出力から式を順に取り出し、取り出しに成功した場合にはステップST92に進む。一方、取り出しに失敗した場合にはそのままこの処理を終了する。ステップST92においては、定型文書入力画面記述・出力文書構造対応表35を参照し、取り出した式の左辺（=の左側の部分）に対するパス表現を得る。次いでステップST93において、その得られたパス表現に対応する出力文書構造木中の枝を探し、ステップS

T94において、取り出した式の右辺（=の右側の部分）の文字列を出力文書構造木中の枝に挿入して、一連の処理を終了する。

【0108】例えば、ステップST91において取り出される式が、図40に点線枠91で囲んで示したものであり、かつ定型文書入力画面記述・出力文書構造対応表35が図41に例示したものである場合には、ステップST92において、パス表現として<購入伺い書>、<氏名>が選択され、ステップST93において、出力文書構造木の枝として図42に点線枠94で示したものが選択され、さらにステップST94において、その出力文書構造木の枝に「森口 太郎」が挿入される。

【0109】以上のように、この実施の形態6によれば、定型文書入力手段34によって作成された、図39に示す定型文書入力画面中の入力枠をユーザが順次うめてゆくことにより、図2に例示したものと同様の出力文書を生成することができ、このように、出力文書構造木生成手段36にて定型文書入力手段34の出力から出力文書構造木を生成することが可能となるため、定型文書の入力画面を利用して構造化文書を編集することができる効果がある。

【0110】実施の形態7、上記実施の形態6においては、定型文書入力画面記述・出力文書構造対応表35を事前に人手で準備する場合について説明したが、それを当該構造化文書処理装置で生成することによって、定型文書の入力画面を利用した構造化文書の編集の手間を削減することも可能である。図44はそのようなこの発明の実施の形態7による構造化文書処理装置の構成を示すブロック図であり、実施の形態6と同様または相当する部分については、図37と同一の符号を付しその説明を省略する。図において、37は対応表代入手段であり、文書構造操作命令解釈実行手段16から呼び出され、文書構造操作命令書15に記述された操作命令にしたがって、定型文書入力画面における入力枠と出力文書構造木中の構造要素識別子の対応関係に関する情報を定型文書入力画面記述・出力文書構造対応表35に代入するものである。

【0111】図45はこの実施の形態7における定型文書入力画面記述書33（出力文書24と同じもの）を生成するために必要な情報を含んだ入力文書の一例を示す説明図であり、購入伺い書テンプレートについて例示したものである。

【0112】図46はこの実施の形態7における文書構造操作命令書15の一例を示す説明図であり、図45に示した出力文書24を入力文書として受理し、図46に示す定型文書入力画面記述・出力文書構造対応表35と、図45に示す定型文書入力画面記述書33を生成するために必要な文書構造の操作命令が記載されている場合について例示している。図において、点線枠95で囲まれた部分是对应表識別子とその対応表識別子の対応表

を記述しており、対応表識別子 T は「定型文書入力画面記述・出力文書構造対応表」であることを示している。また、点線枠 9 6 で囲まれた部分は対応表代入文であり、この対応表代入文の右辺（：＝の右側の部分）を左辺（：＝の左側の部分）の対応表識別子 T が指し示す「定型文書入力画面記述・出力文書構造対応表」に代入することを示している。

【0113】次に動作について説明する。ここで、図 4 7 はこの実施の形態 7 における構造化文書処理装置の全体の動作を示すフローチャートである。なお、基本的な動作は実施の形態 1 において説明したものと同等であるためここではその説明は省略し、それとは異なる部分を中心に説明する。ステップ S T 5 において、取り出した操作命令の命令タイプを判定し、その命令タイプに応じて、構造要素抽出手段 1 7、条件判定手段 1 8、繰り返し構造処理手段 1 9、構造要素代入手段 2 0、対応表代入手段 3 7 の中の必要なものを呼び出して、取り出された操作命令を解釈実行させる。当該操作命令の解釈実行が終了した後はステップ S T 4 に戻り、文書構造操作命令書 1 5 から次の操作命令を取り出す。ステップ S T 5 における判定の結果、命令タイプが対応表代入文であった場合には対応表代入手段 3 7 を呼び出し、ステップ S T 1 4 において対応表代入文を解釈実行する。

【0114】図 4 8 はこのステップ S T 1 4 における文書構造操作命令書 1 5 中の対応表代入文を解釈実行する動作を示すフローチャートである。文書構造操作命令解釈実行手段 1 6 に呼び出された対応表代入手段 3 7 は、まずステップ S T 1 0 1 において、対応表代入文の右辺から入力枠識別子と構造要素識別子を取り出し、次いでステップ S T 1 0 2 において、その取り出した結果を定型文書入力画面記述・出力文書構造対応表 3 5 に代入して、処理を図 4 7 のステップ S T 4 に戻す。例えば、入力文書 1 1 が図 4 5 に例示したものであり、かつ対応表代入文が図 4 6 の点線枠 9 6 で囲んだものである場合には、ステップ S T 1 0 1 において、入力枠識別子として 1 が、そして構造要素識別子として＜購入伺い書＞がそれぞれ取り出され、それをステップ S T 1 0 2 において定型文書入力画面記述・出力文書構造対応表 3 5 に代入した結果が、図 4 1 に点線枠 9 2 で囲んで示した部分にあたる。

【0115】以上のように、この実施の形態 7 によれば、図 4 5 に示した文書を入力文書 1 1 として受理し、図 4 1 に示した定型文書入力画面記述・出力文書構造対応表 3 5 と図 3 8 に示した定型文書入力画面記述書 3 3 を生成することができ、このように、対応表代入文を解釈実行する対応表代入手段 3 7 により、定型文書の入力画面を利用した構造化文書の編集の際に必要となる定型文書入力画面記述書 3 3 と定型文書入力画面記述・出力文書構造対応表 3 5 を、当該構造化文書処理装置で生成することが可能となるため、定型文書の入力画面を利用

した構造化文書の編集の手間を削減することができる効果がある。

【0116】実施の形態 8、上記各実施の形態においては、文書構造操作命令書 1 5 の編集については特に言及していなかったが、入力文書文書型定義書 1 2 と出力文書文書型定義書 2 2 の木構造を表示し、その構造木を操作することによって、文書構造操作命令書 1 5 の編集を行なうことも可能である。図 4 9 はそのようなこの発明の実施の形態 8 による構造化文書処理装置の構成を示すブロック図であり、実施の形態 1 と同様または相当する部分については、図 1 と同一の符号を付しその説明を省略する。図において、3 8 は入出力文書構造対応付け手段であり、入力文書文書型定義書 1 2 と出力文書文書型定義書 2 2 とを読み込んで、それぞれの論理構造を表現する木構造を画面上に表示し、ポインティングデバイスを用いたユーザの操作による両者の木構造の構造要素識別子間の対応付けを受理して、文書構造操作命令書 1 5 を作成するものである。

【0117】図 5 0 はこの実施の形態 8 における入出力文書構造対応付け手段 3 8 が、入力文書文書型定義書 1 2 と出力文書文書型定義書 2 2 を読み込み、木構造で表示することによって得られる入出力文書構造対応付けインタフェースの一例を示す説明図である。図において、点線枠 1 0 1 で囲まれた部分は図 4 に示した入力文書文書型定義書 1 2 の論理構造を表現する木構造であり、点線枠 1 0 2 で囲まれた部分は図 5 に示した出力文書文書型定義書 2 2 の論理構造を表現する木構造である。点線枠 1 0 3 で囲まれた部分は上記木構造（1 0 1）の枝であり、点線枠 1 0 4 で囲まれた部分は上記木構造（1 0 2）の枝である。1 0 5 および 1 0 6 はポインティングデバイスによってその位置が制御されるカーソルである。点線枠 1 0 7 で囲まれた部分は上記カーソル 1 0 5 を用いて木構造（1 0 1）の枝（1 0 3）を指定することにより得られる、入力文書 1 1 の文書構造木のパス表現であり、点線枠 1 0 8 で囲まれた部分は上記カーソル 1 0 6 を用いて木構造（1 0 2）の枝（1 0 4）を指定することにより得られる、出力文書 2 4 の文書構造木のパス表現である。

【0118】図 4 5 は入出力文書構造対応付け手段 3 8 によって得られる、この実施の形態 8 における文書構造操作命令書 1 5 の一例を示す説明図である。図において、点線枠 1 0 9 で囲まれた部分は入力文書 1 1 に対して生成された構造要素抽出文であり、点線枠 1 1 0 で囲まれた部分は出力文書 2 4 に対して生成された構造要素代入文である。

【0119】次に動作について説明する。ここで、図 5 2 は上記入出力文書構造対応付け手段 3 8 の動作を示すフローチャートである。まずステップ S T 1 1 1 において、入力文書文書型定義書 1 2 と出力文書文書型定義書 2 2 とを読み込んで、それぞれの文書型定義の論理構造

を表現する構造木（以下、文書型定義構造木という）を画面上に表示する。次いでステップST112において、入力文書11の文書型定義構造木中の枝に対するポインティングデバイスによるユーザからの指定を受理する。次にステップST113において、その文書型定義構造木中の枝からパス表現の同定を行なった後、ステップST114において、入力文書11に対する構造要素抽出文を生成する。次いでステップST115において、出力文書24の文書型定義構造木中の枝に対するポインティングデバイスによるユーザからの指定を受理する。次にステップST116において、その文書型定義構造木中の枝からパス表現の同定を行なった後、ステップST117において、出力文書24に対する構造要素代入文を生成する。

【0120】例えば、入力文書文書型定義書12が図4に例示したものであり、かつ出力文書文書型定義書22が図5に例示したものである場合には、ステップST111において、図50の点線枠101および102で囲まれた部分に示された木構造が画面表示される。さらに、ユーザがステップST112でポインティングデバイスを用いてカーソル105を移動させ、図50の木構造（101）の枝（103）を指定した場合、ステップST113では入力文書11の文書型定義構造木として、図50に点線枠107で囲んで示したパス表現を得る。そして、ステップST114にて図51に点線枠109で囲んで示した入力文書11に対する構造要素抽出文が生成される。また、ユーザがステップST115において、引き続きポインティングデバイスを用いてカーソル106を移動させ、図50に示した木構造（102）の枝（104）指定した場合、ステップST116では出力文書24の文書型定義構造木として、図50に点線枠108で囲んで示したパス表現を得る。そして、ステップST117にて図51に点線枠110で囲んで示した出力文書24に対する構造要素代入文が生成される。

【0121】以上のように、この実施の形態8によれば、図4および図5に示した入力文書文書型定義書12と出力文書文書型定義書22を読み込み、入出力構造対応付けのためのインタフェースを生成、表示して、ポインティングデバイスによる木構造中の枝の指定を受け付けることにより、図51に示す文書構造操作命令書15を生成することができ、このように、インタフェース画面上に表示された文書型定義構造木をポインティングデバイスを用いて操作することにより、文書構造操作命令書15を編集することが可能となるため、文書構造操作命令書15の編集の手間を削減することができる効果がある。

【0122】

【発明の効果】以上のように、請求項1記載の発明によれば、文書構造操作命令解釈実行手段が文書構造操作命

令書を読み込み、そこに記述された操作命令を順番に取り出して解釈実行することで、文書構造解析手段が入力文書の論理構造を解析して作成した入力文書構造木より出力文書構造木を生成し、構造化文書生成手段にてその出力文書構造木から出力文書を生成するように構成したので、高度な構造化文書の処理を実現することが可能となり、構造化文書処理プログラムの作成の手間を削減することができる構造化文書処理装置が得られる効果がある。

【0123】請求項2記載の発明によれば、文書構造操作命令書を読み込んだ文書構造操作命令解釈実行手段が、その文書構造操作命令書より順次取り出した操作命令のタイプに応じて、構造要素識別子の指定する文書内容を入力文書構造木から抽出する構造要素抽出手段、文書処理内容を構造要素識別子の指定する文書内容に課せられる条件によって制御する条件判定手段、構造要素識別子の指定する各々の文書内容に対する操作命令を解釈実行する繰り返し構造処理手段、出力文書構造木に構造要素識別子が指定する文書内容を挿入する構造要素代入手段の中から必要なものを選択して、操作命令を解釈実行するように構成したので、複数の構成要素識別子名にまたがる文書内容間の制約条件のチェックや、構成要素識別子の出現の文脈に応じてチェックすべき条件が変わる処理の実行といった、より高度な構造化文書の処理が可能になって、構造化文書に対する文書変換や文書の正しさの検証等の能力を向上させることができ、また、対象とする構造化文書の変更や、構造化文書の操作内容の変更に対しても、文書構造操作命令書のみを変更するだけでよくなるので、構造化文書処理プログラムの作成の手間を削減することが可能になる効果がある。

【0124】請求項3記載の発明によれば、複数の入力文書に対する入力文書構造木を管理する複数入力文書構造木管理手段によって、文書構造解析手段から複数の入力文書構造木を受理し、入力文書識別子と入力文書構造木を対応付けて入力文書構造木バッファに格納するとともに、構造要素抽出手段から入力文書識別子と構造要素抽出命令を受理した場合に、その入力文書識別子に対応する入力文書構造木から構造要素識別子が指定する文書の内容を抽出するように構成したので、一方の文書に記載されている内容と他方の文書に記載されている内容とに矛盾がないかをチェックしたり、複数の文書を合成して一つの文書を作成するといった、より高度な構造化文書処理を実現することが可能になる効果がある。

【0125】請求項4記載の発明によれば、複数の出力文書に対する出力文書構造木を管理する複数出力文書構造木管理手段によって、構造要素代入手段から出力文書識別子と構造要素代入命令を受理した場合に、その出力文書識別子に対応する出力文書構造木に構造要素識別子が指定する文書の内容を挿入するように構成したので、文書中の構造要素が満たすべき条件に応じて文書を分割

したり、複数の入力文書から複数の出力文書を生成するといった、より高度な構造化文書処理を実現することが可能になる効果がある。

【0126】請求項5記載の発明によれば、文書構造操作命令書を読み込んだ文書構造操作命令解釈実行手段が、その文書構造操作命令書より順次取り出した操作命令のタイプに応じて、構造要素抽出手段、条件判定手段、繰り返し構造処理手段、構造要素代入手段、さらには辞書ファイルから辞書と呼び出した辞書を参照して入力文書中に含まれる用語の処理を行なう用語処理手段の中から必要なものを選択して、操作命令を解釈実行するように構成したので、入力文書中に含まれる用語について所望の処理を行なうことが可能になり、構造化文書に対する文書変換や文書の正しさの検証等の処理能力を向上させることができる効果がある。

【0127】請求項6記載の発明によれば、辞書ファイルに用語検証に必要な用語と用語に付随する情報を収録した用語辞書を格納し、入力文書中の用語が呼び出された用語辞書に収録されている用語かどうかを用語検証手段にて検証するように構成したので、入力文書中に含まれる用語に関する誤りの検出や、誤った用語の修正を行なうことが可能となり、構造化文書に対する文書の正しさについての検証能力を向上させることができる効果がある。

【0128】請求項7記載の発明によれば、辞書ファイルに変換前の用語と変換後の用語との対応関係を収録した変換辞書を格納し、用語変換手段が呼び出された変換辞書を参照して、入力文書中に含まれる用語を当該用語に対応付けられている別の用語に変換するように構成したので、使用している用語の表記を変更した文書を生成することが可能になる効果がある。

【0129】請求項8記載の発明によれば、定型文書入力手段において、定型文書の入力用画面のレイアウトを記述した定型文書入力画面記述書を読み込んで定型文書の入力用画面を作成した後、ユーザの定型文書入力を受け取ってその結果を出力文書構造木生成手段に送り、出力文書構造木生成手段において、定型文書入力画面記述書の記述と出力文書の論理構造の対応関係を記述した定型文書入力画面記述・出力文書構造木対応表を読み込んで、定型文書入力手段の出力と定型文書入力画面記述・出力文書構造木対応表の内容から出力文書構造木を生成し、それを出力文書構造木バッファに格納するように構成したので、定型文書の入力画面を利用して構造化文書を編集することが可能となり、構造化文書の作成を容易化することができる効果がある。

【0130】請求項9記載の発明によれば、文書構造操作命令書を読み込んだ文書構造操作命令解釈実行手段が、その文書構造操作命令書より順次取り出した操作命令のタイプに応じて、構造要素抽出手段、条件判定手段、繰り返し構造処理手段、構造要素代入手段、さらに

は定型文書入力画面における入力枠と出力文書構造木中の構造要素識別子の対応関係に関する情報を定型文書入力画面記述・出力文書構造木対応表に代入して、定型文書入力画面記述・出力文書構造木対応表を作成する対応表代入手段の中から必要なものを選択して、操作命令を解釈実行するように構成したので、定型文書の入力画面を利用した構造化文書の編集の際に必要な定型文書入力画面記述書と定型文書入力画面記述・出力文書構造対応表を生成することができるようになり、定型文書の入力画面を利用した構造化文書の編集の手間を削減することが可能となって、構造化文書の作成をより効率化することができる効果がある。

【0131】請求項10記載の発明によれば、入出力文書構造対応付け手段にて入力文書文書型定義書と出力文書文書型定義書を読み込み、それぞれの論理構造を表現する木構造を表示し、それら両者の木構造の構造要素識別子間の対応付けを受け取って文書構造操作命令書を作成するように構成したので、画面上に表示された文書型定義構造木をポインティングデバイスなどを用いて操作することにより、文書構造操作命令書を編集することが可能となり、文書構造操作命令書の編集の手間を軽減することができ、構造化文書の作成をより効率的に行なうことができる効果がある。

【図面の簡単な説明】

【図1】 この発明の実施の形態1による構造化文書処理装置の構成を示すブロック図である。

【図2】 この発明の実施の形態1における入力文書の一例を示す説明図である。

【図3】 この発明の実施の形態1における出力文書の一例を示す説明図である。

【図4】 この発明の実施の形態1における入力文書文書型定義書の一例を示す説明図である。

【図5】 この発明の実施の形態1における出力文書文書型定義書の一例を示す説明図である。

【図6】 この発明の実施の形態1における入力文書構造木の一例を示す説明図である。

【図7】 この発明の実施の形態1における文書構造操作命令書の一例を示す説明図である。

【図8】 この発明の実施の形態1における出力文書構造木の一例を示す説明図である。

【図9】 この発明の実施の形態1における出力文書の文書型定義構造木の一例を示す説明図である。

【図10】 この発明の実施の形態1における構造化文書処理装置の全体動作を示すフローチャートである。

【図11】 この発明の実施の形態1における構造要素抽出手段の動作を示すフローチャートである。

【図12】 この発明の実施の形態1における構造要素代入手段の動作を示すフローチャートである。

【図13】 この発明の実施の形態1における構造化文書生成手段の動作を示すフローチャートである。

【図 1 4】 この発明の実施の形態 2 による構造化文書処理装置の構成を示すブロック図である。

【図 1 5】 この発明の実施の形態 2 における入力文書の一例を示す説明図である。

【図 1 6】 この発明の実施の形態 2 における入力文書識別子・入力文書構造木対応表の一例を示す説明図である。

【図 1 7】 この発明の実施の形態 2 における文書構造操作命令書の一例を示す説明図である。

【図 1 8】 この発明の実施の形態 2 における構造要素抽出手段の動作を示すフローチャートである。

【図 1 9】 この発明の実施の形態 3 による構造化文書処理装置の構成を示すブロック図である。

【図 2 0】 この発明の実施の形態 3 における出力文書の一例を示す説明図である。

【図 2 1】 この発明の実施の形態 3 における出力文書識別子・出力文書構造木対応表の一例を示す説明図である。

【図 2 2】 この発明の実施の形態 3 における文書構造操作命令書の一例を示す説明図である。

【図 2 3】 この発明の実施の形態 3 における構造要素代入手段の動作を示すフローチャートである。

【図 2 4】 この発明の実施の形態 4 による構造化文書処理装置の構成を示すブロック図である。

【図 2 5】 この発明の実施の形態 4 における入力文書の一例を示す説明図である。

【図 2 6】 この発明の実施の形態 4 における出力文書の一例を示す説明図である。

【図 2 7】 この発明の実施の形態 4 における用語辞書の一例を示す説明図である。

【図 2 8】 この発明の実施の形態 4 における文書構造操作命令書の一例を示す説明図である。

【図 2 9】 この発明の実施の形態 4 における構造化文書処理装置の全体動作を示すフローチャートである。

【図 3 0】 この発明の実施の形態 4 における用語検証手段動作を示すフローチャートである。

【図 3 1】 この発明の実施の形態 5 による構造化文書処理装置の構成を示すブロック図である。

【図 3 2】 この発明の実施の形態 5 における入力文書の一例を示す説明図である。

【図 3 3】 この発明の実施の形態 5 における出力文書の一例を示す説明図である。

【図 3 4】 この発明の実施の形態 5 における変換辞書の一例を示す説明図である。

【図 3 5】 この発明の実施の形態 5 における文書構造操作命令書の一例を示す説明図である。

【図 3 6】 この発明の実施の形態 5 における用語変換手段の動作を示すフローチャートである。

【図 3 7】 この発明の実施の形態 6 による構造化文書処理装置の構成を示すブロック図である。

【図 3 8】 この発明の実施の形態 6 における定型文書入力画面記述書の一例を示す説明図である。

【図 3 9】 この発明の実施の形態 6 における定型文書入力画面の一例を示す説明図である。

【図 4 0】 この発明の実施の形態 6 における定型文書入力手段の出力の一例を示す説明図である。

【図 4 1】 この発明の実施の形態 6 における定型文書入力画面記述・出力文書構造記述対応表の一例を示す説明図である。

【図 4 2】 この発明の実施の形態 6 における出力文書構造木生成手段によって生成される出力文書構造木の一例を示す説明図である。

【図 4 3】 この発明の実施の形態 6 における出力文書構造木生成手段の動作を示すフローチャートである。

【図 4 4】 この発明の実施の形態 7 による構造化文書処理装置の構成を示すブロック図である。

【図 4 5】 この発明の実施の形態 7 における入力文書の一例を示す説明図である。

【図 4 6】 この発明の実施の形態 7 における文書構造操作命令書の一例を示す説明図である。

【図 4 7】 この発明の実施の形態 7 における構造化文書処理装置の全体動作を示すフローチャートである。

【図 4 8】 この発明の実施の形態 7 における対応表代入手段の動作を示すフローチャートである。

【図 4 9】 この発明の実施の形態 8 による構造化文書処理装置の構成を示すブロック図である。

【図 5 0】 この発明の実施の形態 8 における入出力文書構造対応付けインタフェースの一例を示す図である。

【図 5 1】 この発明の実施の形態 8 における文書構造操作命令書の一例を示す説明図である。

【図 5 2】 この発明の実施の形態 8 における出力文書構造対応付け手段の動作を示すフローチャートである。

【図 5 3】 従来の構造化文書処理装置の構成を示すブロック図である。

【図 5 4】 従来の構造化文書処理装置におけるチェック項目表の一例を示す説明図である。

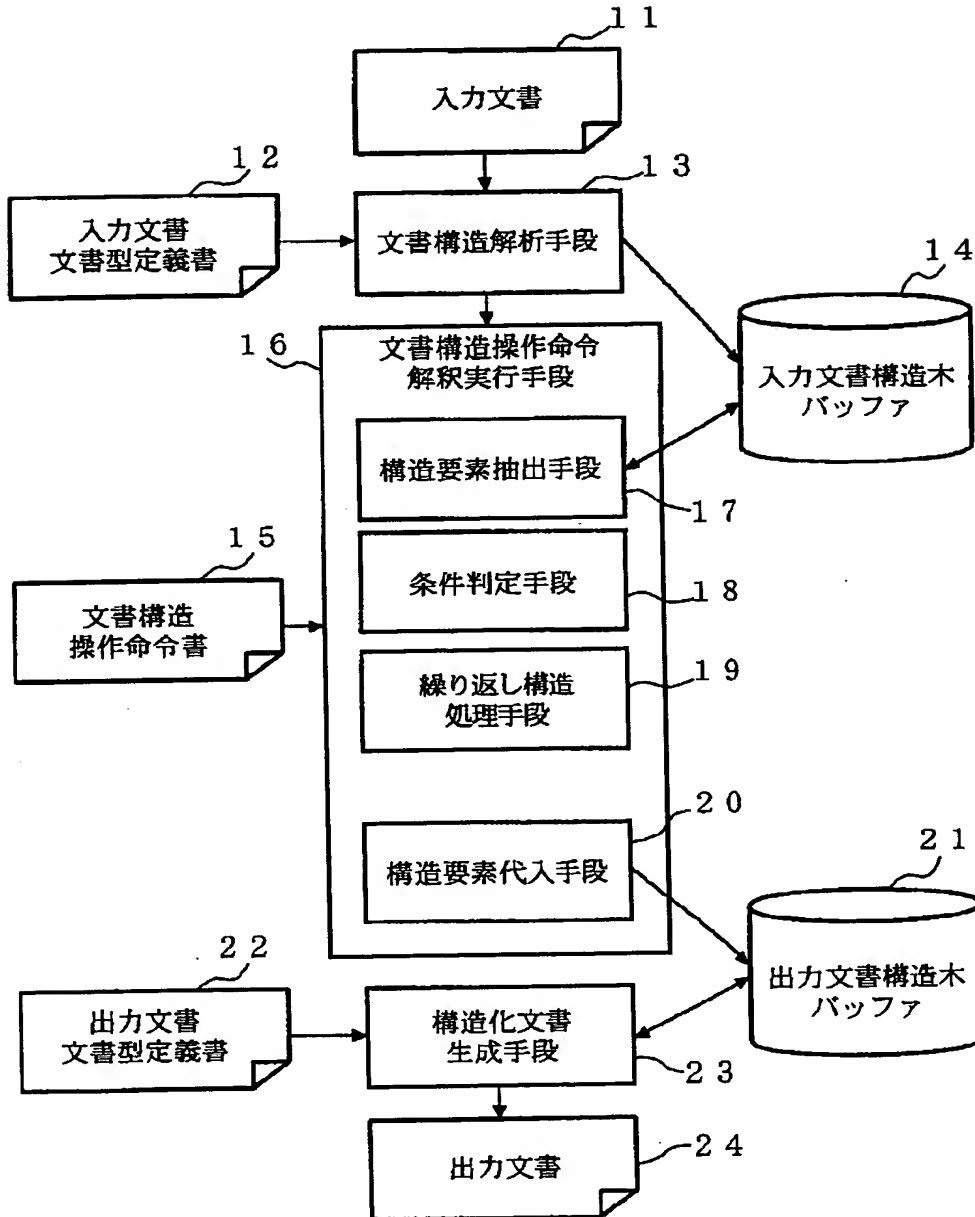
【符号の説明】

- 1 1 入力文書、1 2 入力文書文書型定義書、1 3 文書構造解析手段、1 4 入力文書構造木バッファ、1 5 文書構造操作命令書、1 6 文書構造操作命令解釈実行手段、1 7 構造要素抽出手段、1 8 条件判定手段、1 9 繰り返し構造処理手段、2 0 構造要素代入手段、2 1 出力文書構造木バッファ、2 2 出力文書文書型定義書、2 3 構造化文書生成手段、2 4 出力文書、2 5 複数入力文書構造木管理手段、2 6 複数出力文書構造木管理手段、2 7 用語辞書ファイル（辞書ファイル）、2 8 用語検証手段（用語処理手段）、3 0 変換辞書ファイル（辞書ファイル）、3 1 用語変換手段（用語処理手段）、3 2 変換辞書管理手段、3 3 定型文書入力画面記述書、3 4 定型文書入力手

段、35 定型文書入力画面記述・出力文書構造木対応表、36 出力文書構造木生成手段、37 対応表代入

手段、38 入出力文書構造対応付け手段。

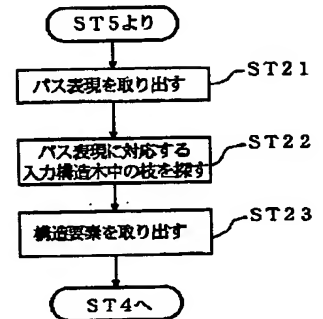
【図1】



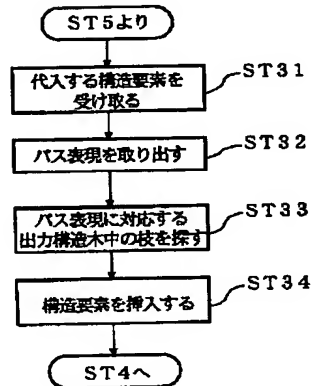
【図5】

<ELEMENT 購入許可書 (許可番号, 検印チェック, 合計チェック, 購入許可品, 見残り)>
 <ELEMENT 検印チェック (*PCDATA)>
 <ELEMENT 合計チェック (*PCDATA)>
 <ELEMENT 許可番号 (*PCDATA)>
 <ELEMENT 購入許可品 (品名*)>
 <ELEMENT 品名 (*PCDATA)>
 <ELEMENT 見残り (社名*)>
 <ELEMENT 社名 (*PCDATA)>

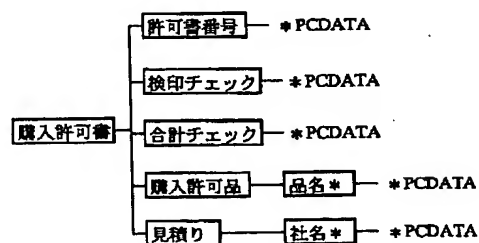
【図11】



【図12】



【図9】



【図2】

<購入伺い書>
 <伺い書番号>314</伺い書番号>
 <氏名> 森口 太郎 </氏名>
 <社員番号> 15926 </社員番号>
 <購入希望品>
 <項目><名称>ワープロA</名称><価格>50000</価格></項目>
 <項目><名称>プリンタB</名称><価格>300000</価格></項目>
 <項目><名称>表計算ソフトC</名称><価格>10000</価格></項目>
 </購入希望品>
 <合計>360000</合計>
 <購入希望会社><名称>D社</名称>
 <名称>E社</名称></購入希望会社>
 <部長検印>あり</部長検印>
 </購入伺い書>

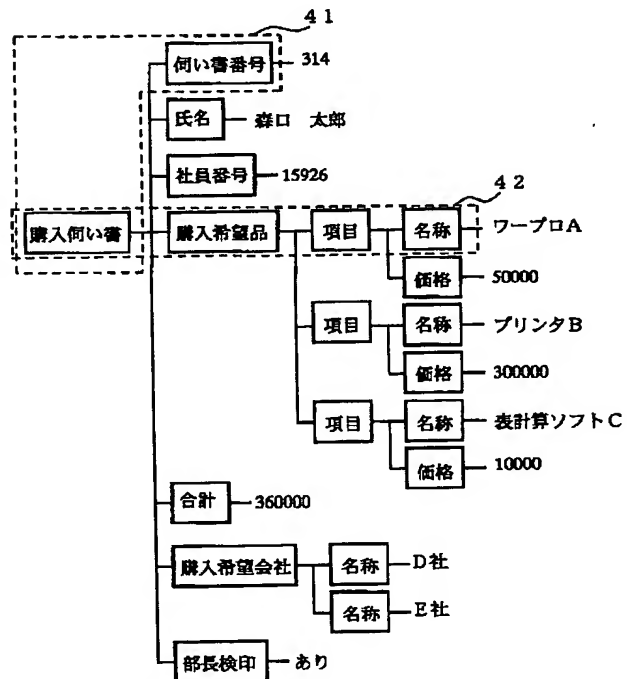
【図3】

<購入許可書>
 <許可書番号>314</許可書番号>
 <検印チェック>成功
 <合計チェック>成功
 <購入許可品>
 <品名>ワープロA</品名>
 <品名>プリンタB</品名>
 <品名>表計算ソフトC</品名>
 </購入許可品>
 <見積り><社名>D社</社名>
 <社名>E社</社名></見積り>
 </購入許可書>

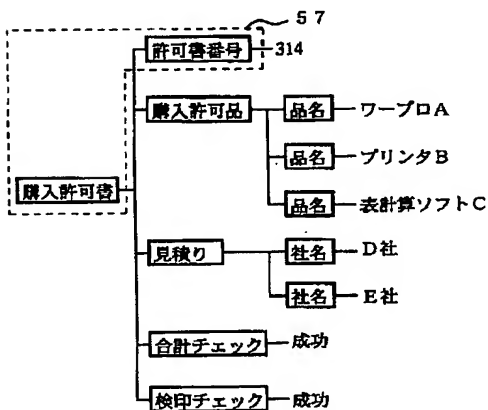
【図4】

<ELEMENT 購入伺い書 (伺い書番号, 氏名, 社員番号, 購入希望品, 合計, 購入希望会社, 部長検印)>
 <ELEMENT 伺い書番号 (*PCDATA)>
 <ELEMENT 氏名 (*PCDATA)>
 <ELEMENT 社員番号 (*PCDATA)>
 <ELEMENT 購入希望品 (項目*)>
 <ELEMENT 項目 (名称, 価格)>
 <ELEMENT 名称 (*PCDATA)>
 <ELEMENT 価格 (*PCDATA)>
 <ELEMENT 合計 (*PCDATA)>
 <ELEMENT 購入希望会社 (名称*)>
 <ELEMENT 部長検印 (*PCDATA)>

【図6】



【図8】

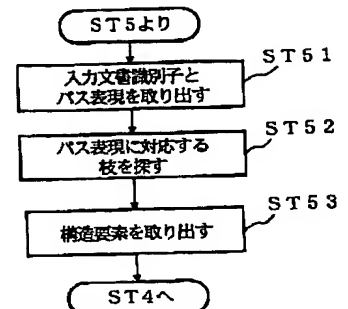


【図15】

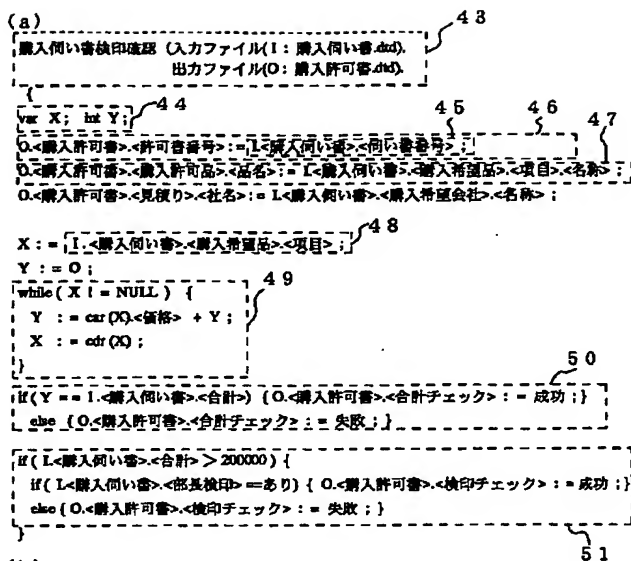
<従業員名簿>
 <従業員><氏名> 森口 太郎 </氏名>
 <社員番号> 15926 </社員番号>
 <従業員><氏名> 今村 次郎 </氏名>
 <社員番号> 53589 </社員番号>
 <従業員><氏名> 藤井 三郎 </氏名>
 <社員番号> 79323 </社員番号>

 </従業員名簿>

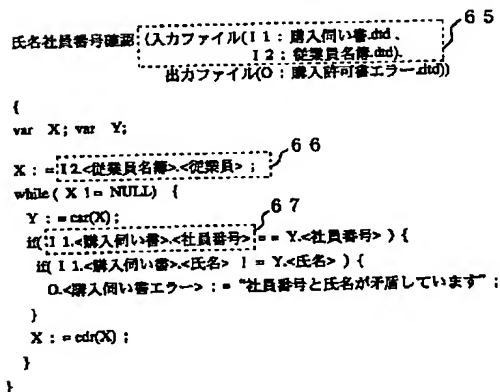
【図18】



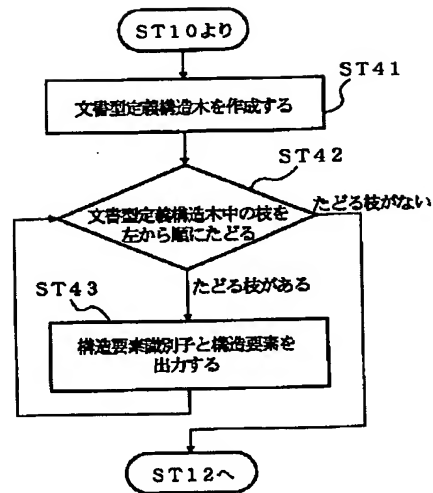
【図7】



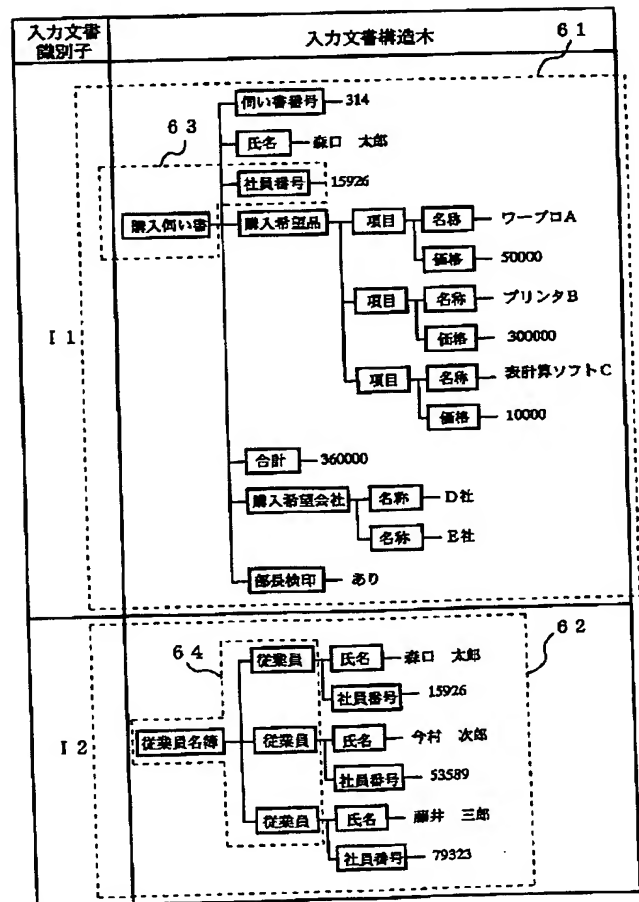
【図17】



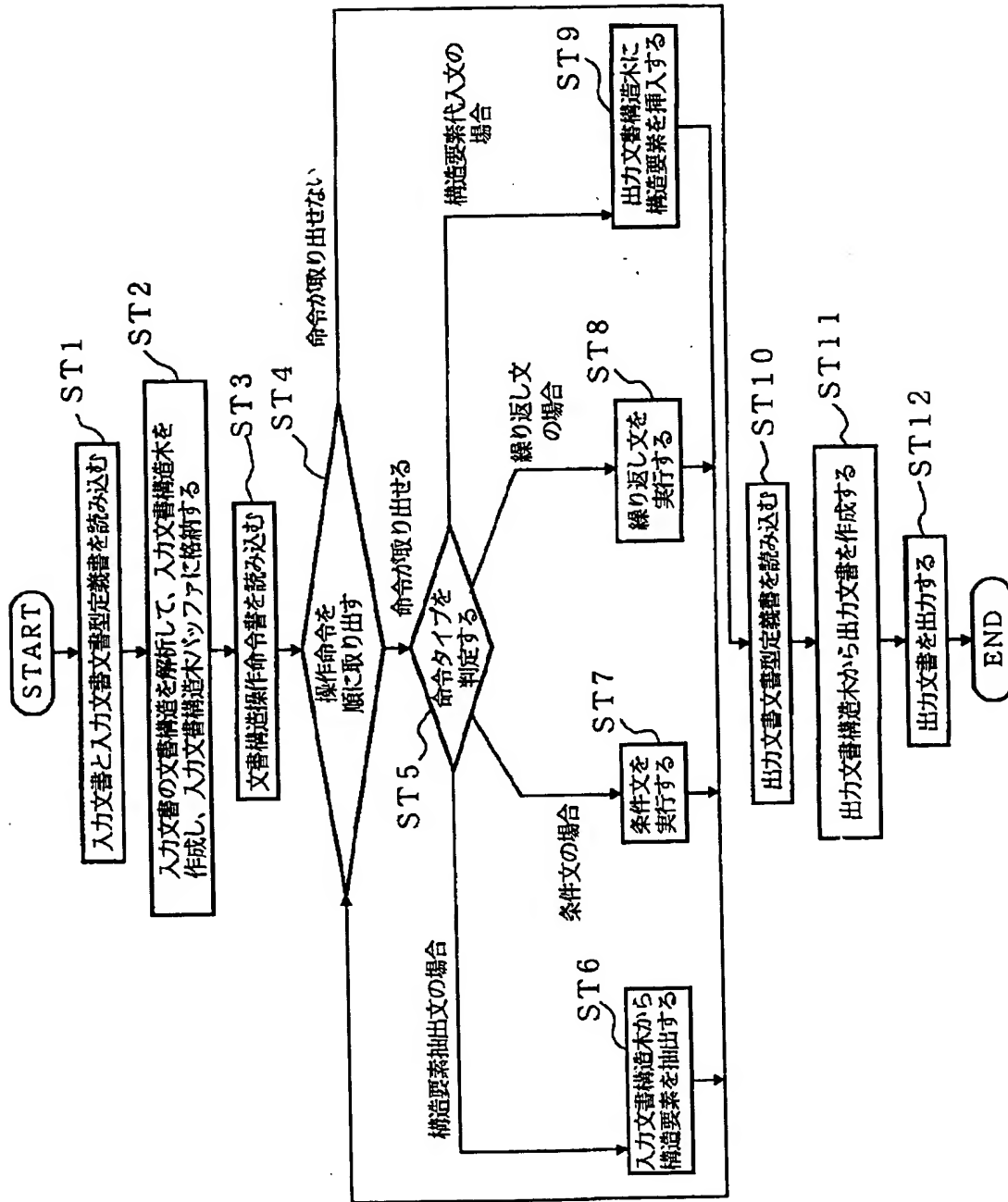
【図13】



【図16】

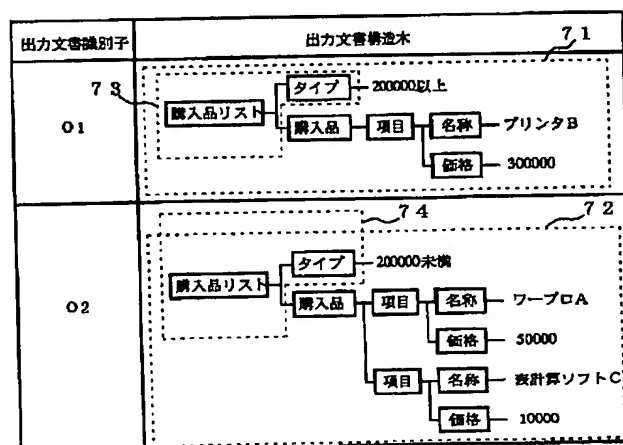
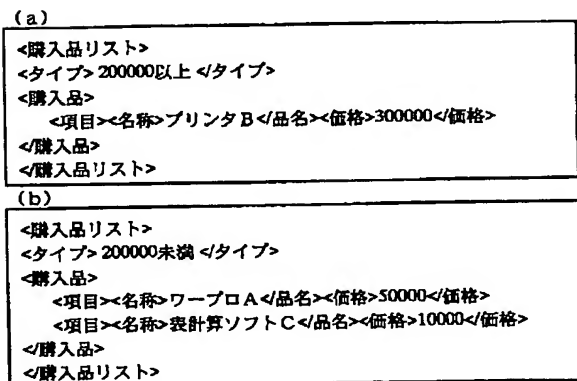


【図10】

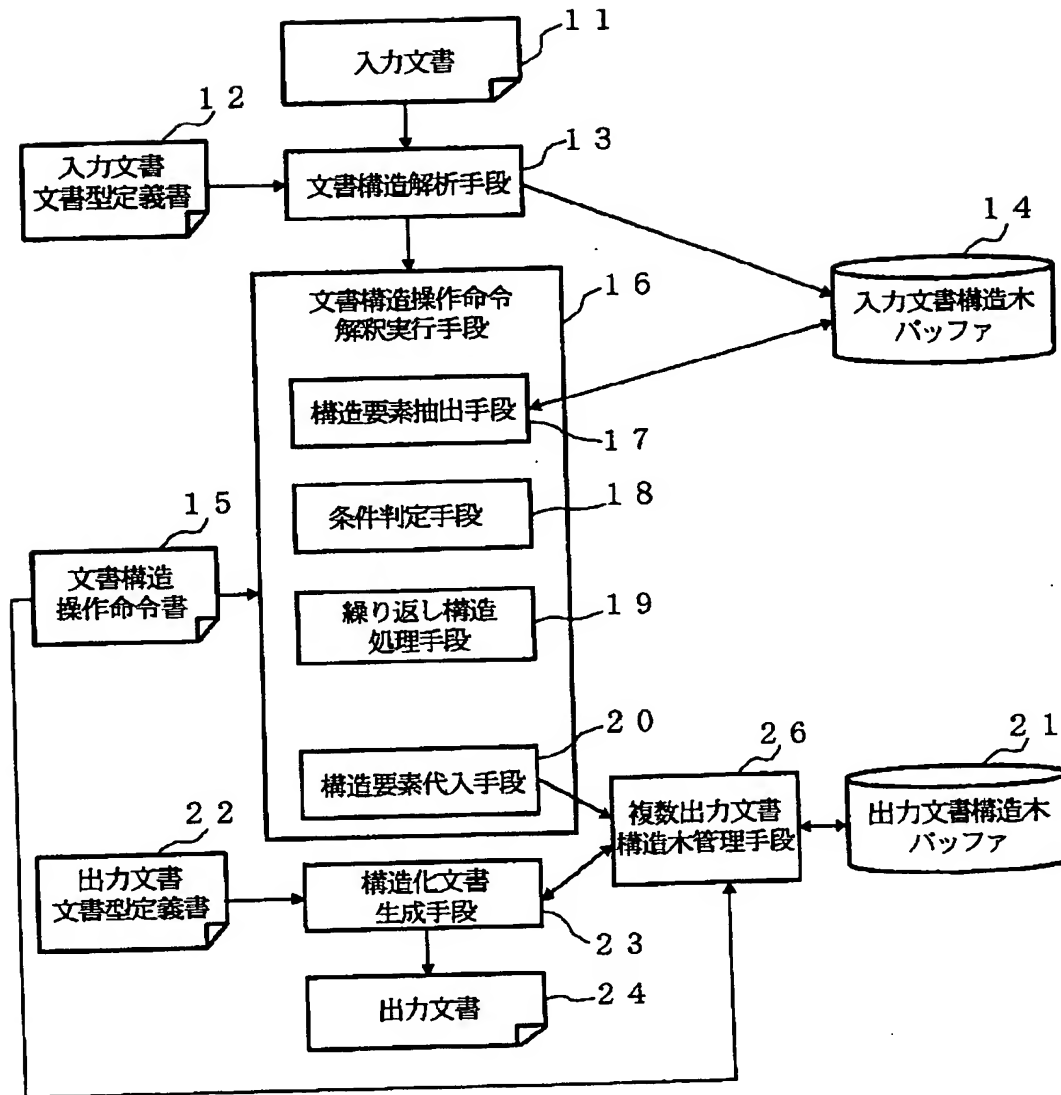


```
graph TD
    11[1.1 入力文書] --> 13[1.3 文書構造解析手段]
    12[1.2 入力文書  
文書型定義書] --> 13
    13 --> 16[1.6 文書構造操作命令  
解釈実行手段]
    13 --> 25[2.5 複数入力文書  
構造木管理手段]
    14[(1.4 入力文書構造木  
バッファ)] <--> 25
    15[1.5 文書構造  
操作命令書] --> 16
    16 --> 17[1.7 構造要素抽出手段]
    16 --> 18[1.8 条件判定手段]
    16 --> 19[1.9 繰り返し構造  
処理手段]
    16 --> 20[2.0 構造要素代入手段]
    20 --> 23[2.3 構造化文書  
生成手段]
    22[2.2 出力文書  
文書型定義書] --> 23
    21[(2.1 出力文書構造木  
バッファ)] --> 23
    23 --> 24[2.4 出力文書]
```

【図 2 1】



【図19】



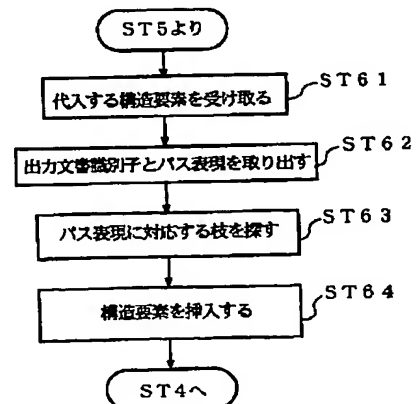
【図22】

```

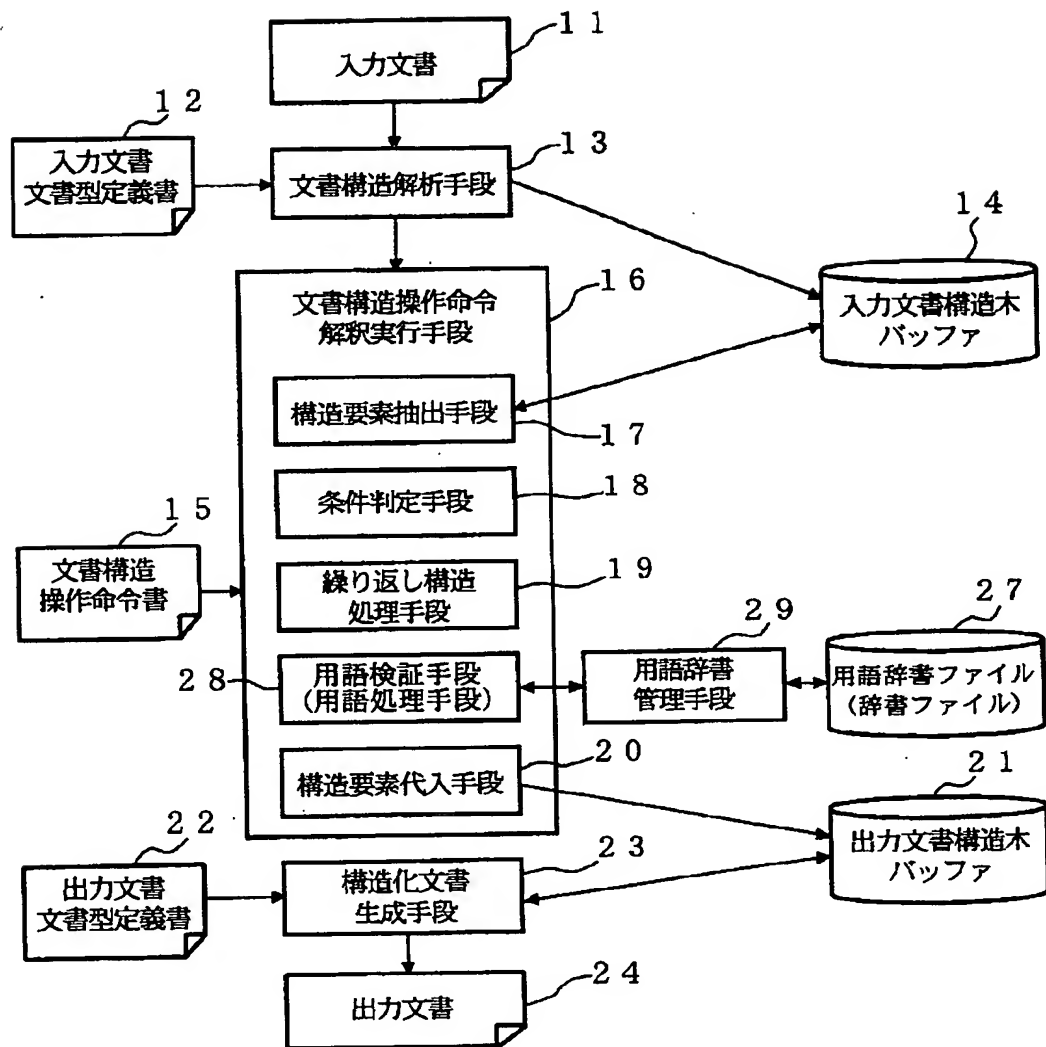
購入品振り分け (入力ファイル(I: 購入許可書.dtd),
                  出力ファイル(O1: 購入品リスト.dtd),
                  O2: 購入品リスト.dtd),
{
  var X;
  O1.<購入許可書>.<購入品リスト> := "200000以上";
  O2.<購入許可書>.<購入品リスト> := "200000未満";

  X := I1.<購入許可書>.<購入品>.<項目>;
  while( X != NULL ) {
    if( car(X).<価格> >= 200000 ) { O1.<購入品>.<項目> := car(X); }
    else{ O2.<購入品>.<項目> := car(X); }
    X := cdr(X);
  }
}
  
```

【図23】



【図24】



【図25】

<参考文献>
 <項目><タイトル>プログラミング言語の試作 <タイトル>
 <雑誌名称>情報処理学会全国大会 <雑誌名称>
 <項目><タイトル>文書検索方式について <タイトル>
 <雑誌名称>情報処理 <雑誌名称>
 </参考文献>

【図26】

<参考文献記述エラー>
 <項目><雑誌名称> 情報処理 <雑誌名称>
 <エラーメッセージ>
 雑誌名辞書にない参考文献が引用されています
 <エラーメッセージ>
 </参考文献記述エラー>

【図32】

<参考文献>
 <項目><タイトル>プログラミング言語の試作 <タイトル>
 <雑誌名称>情報処理学会全国大会 <雑誌名称>
 <項目><タイトル>文書検索方式について <タイトル>
 <雑誌名称>情報処理 <雑誌名称>
 </参考文献>

【図33】

<参考文献>
 <項目><タイトル>プログラミング言語の試作 <タイトル>
 <雑誌名称>情報処理学会全国大会 <雑誌名称>
 <項目><タイトル>文書検索方式について <タイトル>
 <雑誌名称>情報処理学会論文誌 <雑誌名称>
 </参考文献>

【図27】

辞書名：雑誌名称辞書
情報処理学会全国大会 情報処理学会論文誌 情報処理学会研究報告 電子情報通信学会論文誌A 電子情報通信学会論文誌B 電子情報通信学会論文誌C 電子情報通信学会論文誌D-I 電子情報通信学会論文誌D-II

【図28】

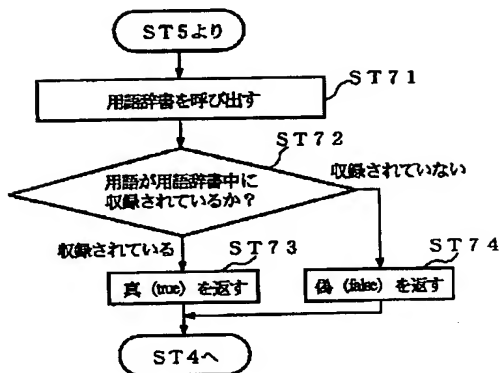
```

参考文献チェック (入力ファイル(I: 参考文献.dtd), 81
                  辞書(D1: 雑誌名称辞書)
                  出力ファイル(O: 参考文献記述エラー.dtd))

{
  var X; var Y;
  X := I.<参考文献>.<項目>; 82
  while (X != NULL) {
    Y := car(X); 83
    if (Y.<雑誌名称> ∈ D1) {
      O.<参考文献記述エラー>.<項目>.<雑誌名> := Y.<雑誌名称>;
      O.<参考文献記述エラー>.<項目>.<メッセージ>
        := "雑誌名称辞書にない参考文献が引用されています";
    }
    X := cdr(X);
  }
}

```

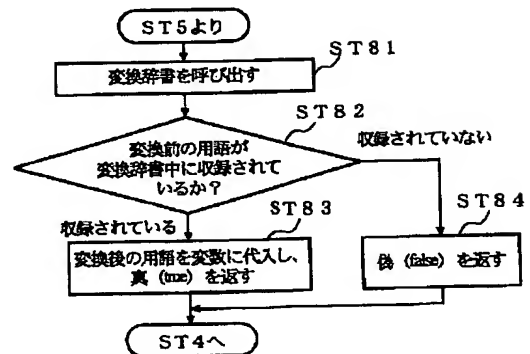
【図30】



【図34】

辞書名：雑誌略称・正式名称対応名辞書	
変換前の用語	変換後の用語
情報全大 情報誌	情報処理学会全国大会 情報処理学会論文誌

【図36】



【図35】

```

参考文献置換 (入力ファイル(I: 参考文献.dtd), 85
                  辞書(D2: 雑誌名称略称・正式名称対応辞書)
                  出力ファイル(O: 参考文献.dtd))

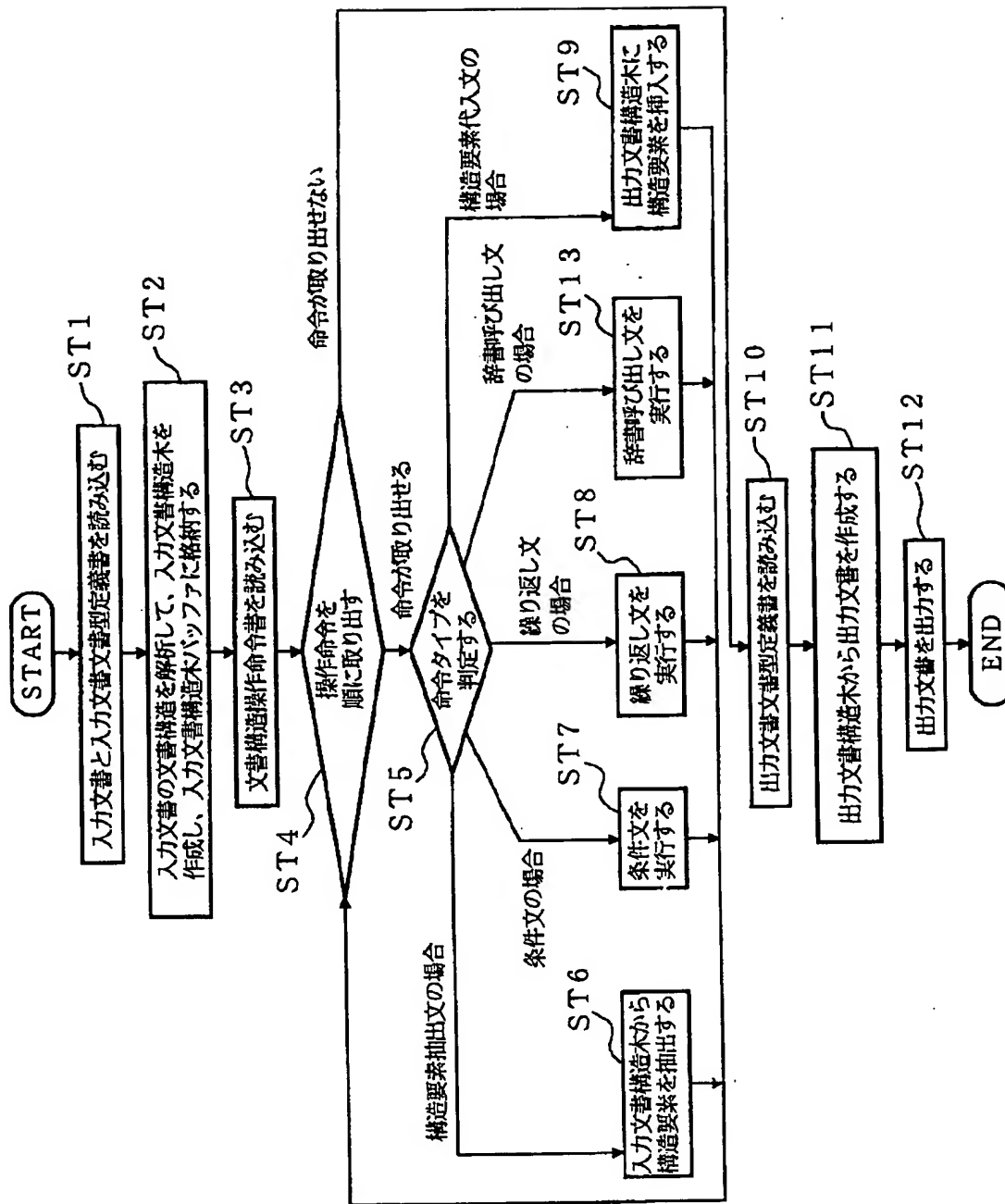
{
  var X; var Y, Z;
  X := I.<参考文献>.<項目>;
  while (X != NULL) {
    Y := car(X); 86
    if (Y.<雑誌名称>, Z := (変換前, 変換後) ∈ D2) { 87
      O.<参考文献>.<項目>.<雑誌名> := Z;
      O.<参考文献>.<項目>.<タイトル> := Y.<参考文献>.<項目>.<タイトル>;
    }
    else {
      O.<参考文献>.<項目>.<雑誌名> := Y.<参考文献>.<項目>.<雑誌名>;
      O.<参考文献>.<項目>.<タイトル> := Y.<参考文献>.<項目>.<タイトル>;
    }
    X := cdr(X);
  }
}

```

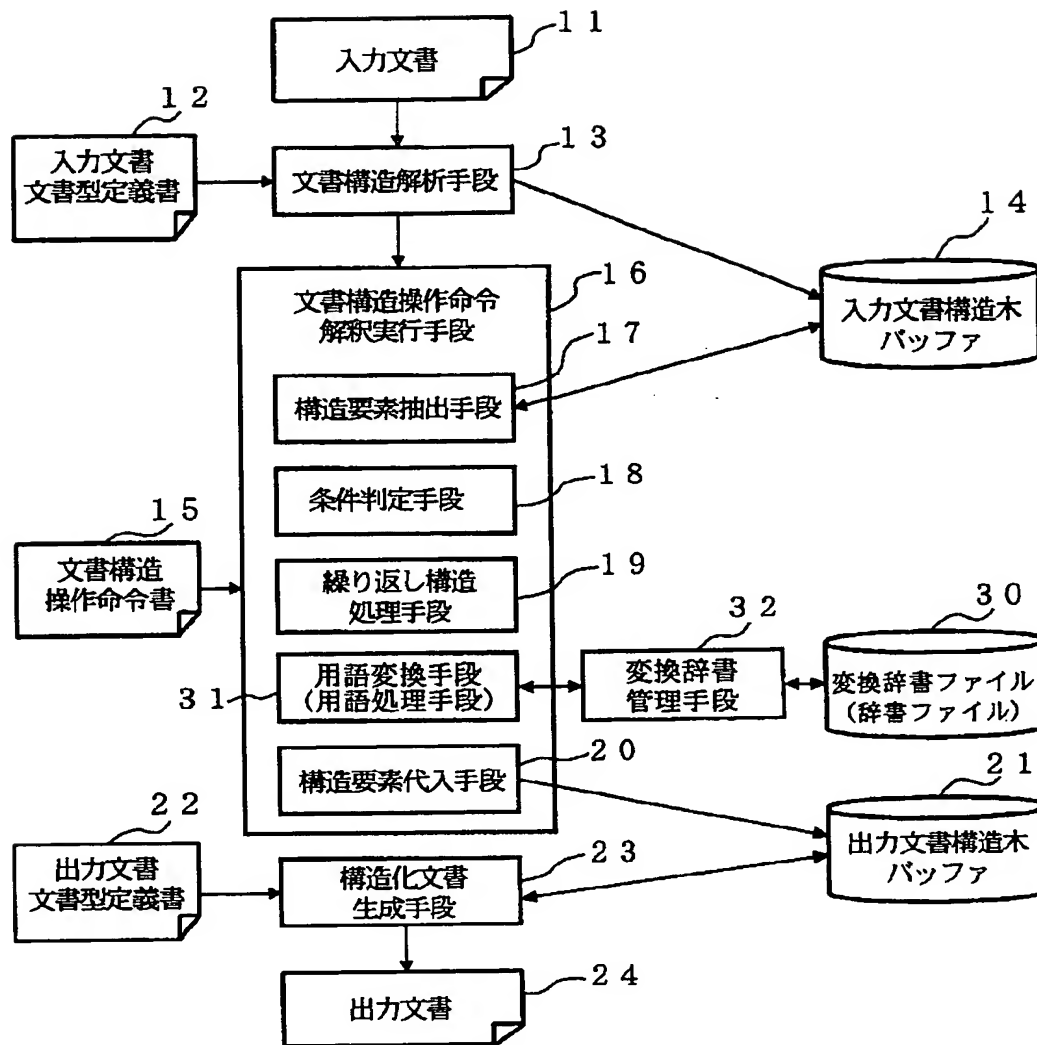
【図39】

氏名	<div> <div>氏名</div> <div>今村 次郎</div> <div>藤井 三郎</div> </div>		
品名	<input type="text"/>	価格	<input type="text"/>
品名	<input type="text"/>	価格	<input type="text"/>
品名	<input type="text"/>	価格	<input type="text"/>

【図29】



【図31】



【図38】

```

<HTML>
<FORM ACTION="cgi-bin/購入問い合わせ入力" METHOD="GET">
<DL>
<DT>氏名
<SELECT NAME=1.2>
<OPTION VALUE="森口 太郎">
<OPTION VALUE="今村 次郎">
<OPTION VALUE="藤井 三郎">
</SELECT>
<DT>
<DT>品名: <INPUT TYPE="TEXT" NAME=1.3.4.5 VALE="">
価格: <INPUT TYPE="TEXT" NAME=1.3.4.6 VALE="">
<DT>品名: <INPUT TYPE="TEXT" NAME=1.3.4.7 VALE="">
価格: <INPUT TYPE="TEXT" NAME=1.3.4.8 VALE="">
<DT>品名: <INPUT TYPE="TEXT" NAME=1.3.4.9 VALE="">
価格: <INPUT TYPE="TEXT" NAME=1.3.4.10 VALE="">
.....
</HTML>

```

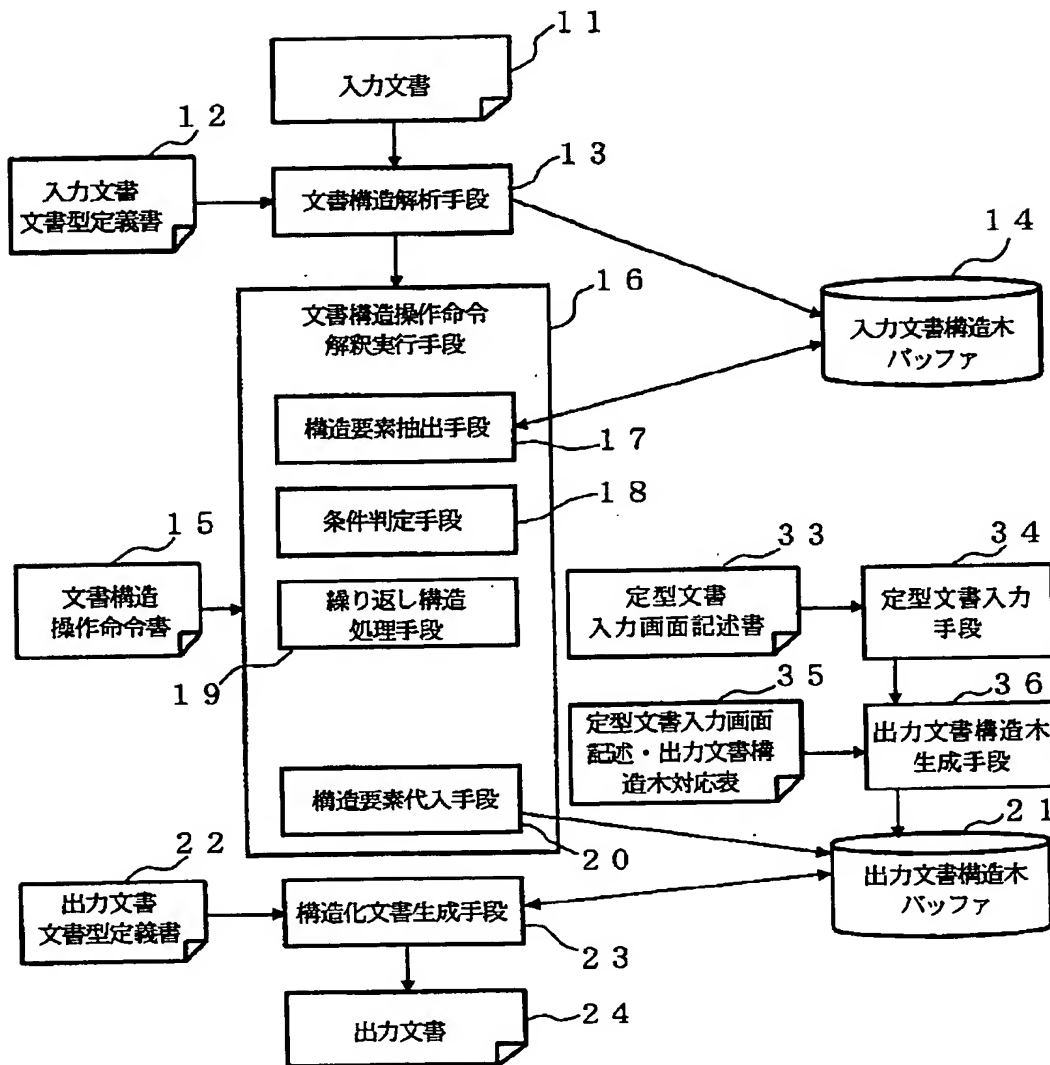
【図40】

91
 &1.2=森口 太郎;&1.3.4.5=ワープロA&1.3.4.6=50000
 &1.3.4.7=プリンタB&1.3.4.8=300000
 &1.3.4.9=表計算ソフトC&1.3.4.10=10000

【図51】

Q.<購入許可書><許可書番号>:=I.<購入問い合わせ><問い合わせ番号>; 109 110
 Q.<購入許可書><購入許可品><品名>:=I.<購入問い合わせ><購入許可品><項目><名称>;
 Q.<購入許可書><見積り><社名>:=I.<購入問い合わせ><購入希望会社><名称>;

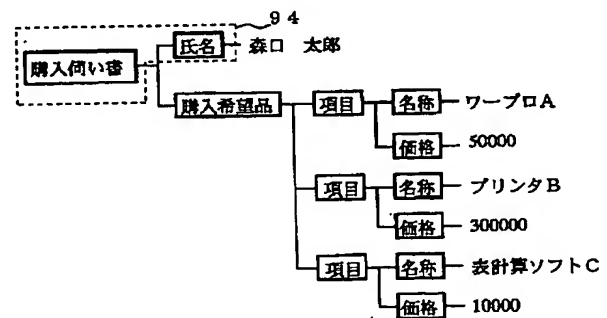
【図37】



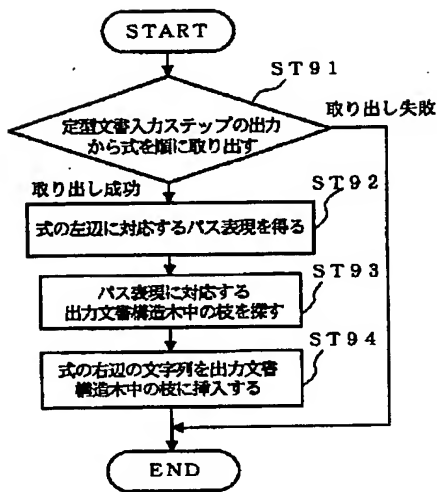
【図41】

入力枠識別子	出力文書構造木の構造要素識別子	
1	<購入問い合わせ>	92
2	<氏名>	93
3	<購入希望書>	
4	<項目>	
5	<名称>	
6	<価格>	
7	<名称>	
8	<価格>	
9	<名称>	
10	<価格>	

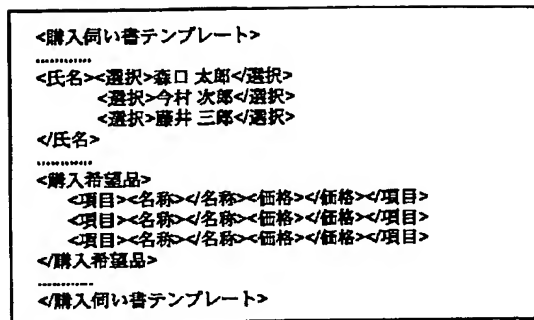
【図42】



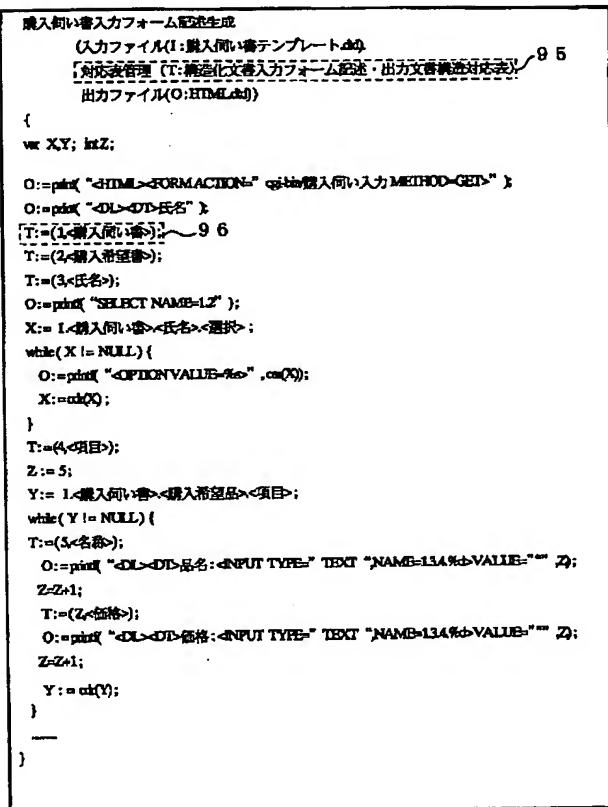
【図43】



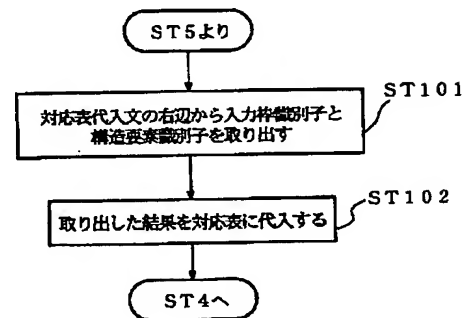
【図45】



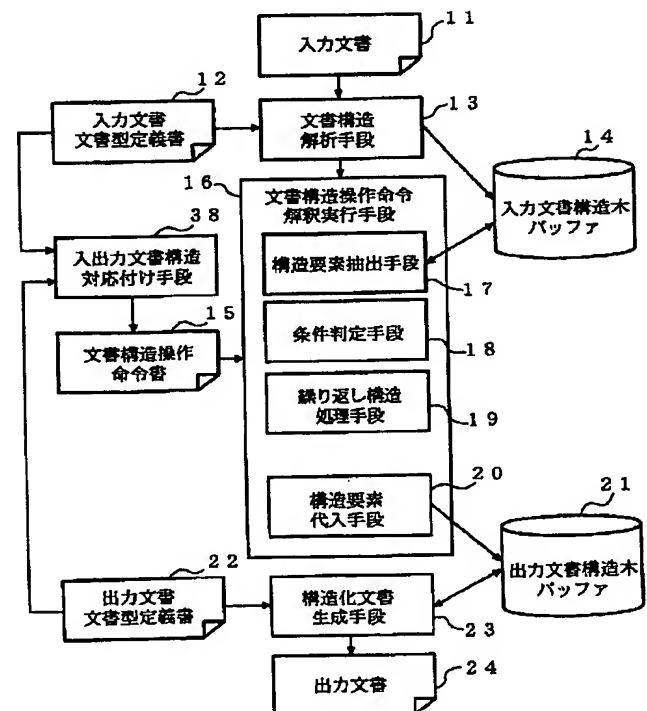
【図46】



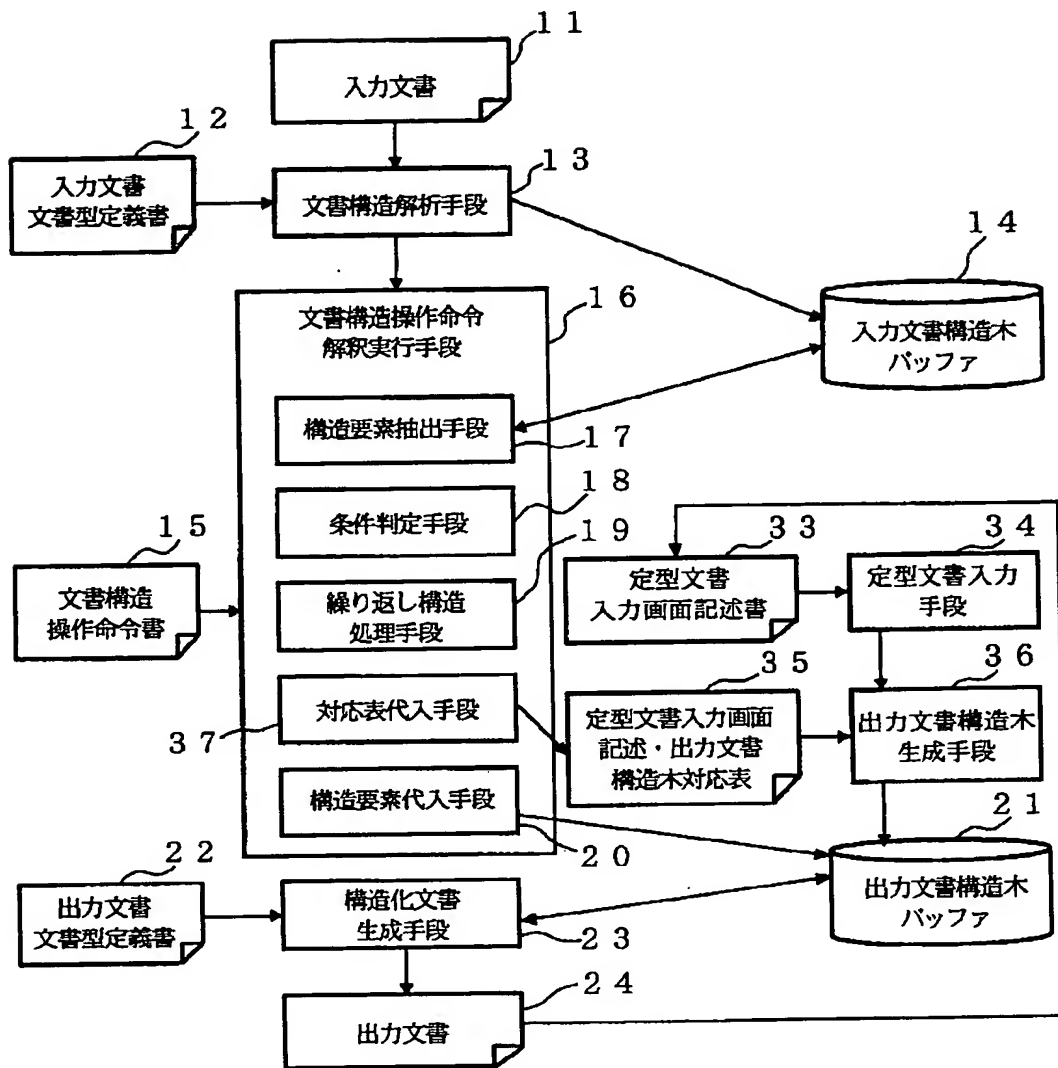
【図48】



【図49】



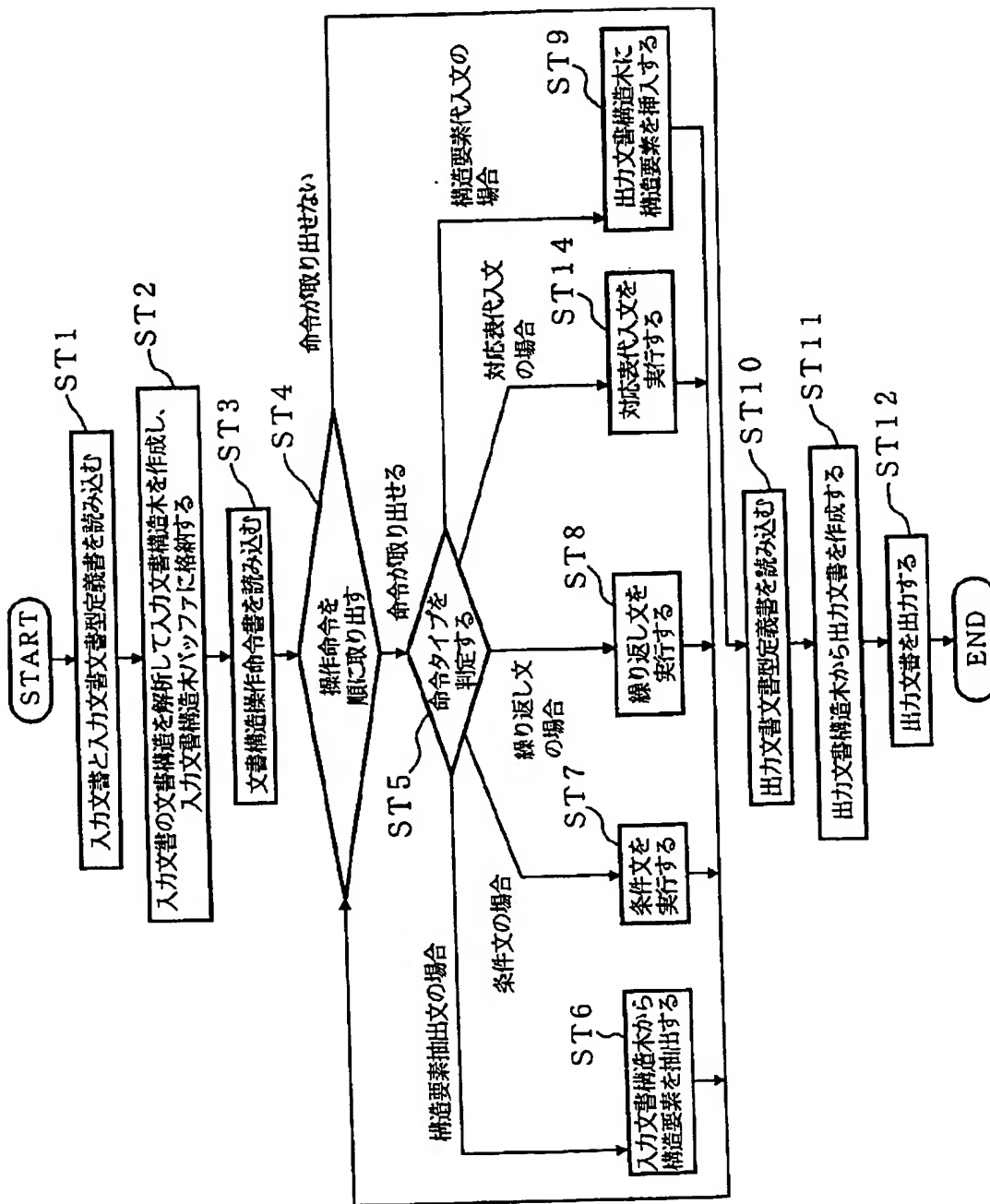
【図44】



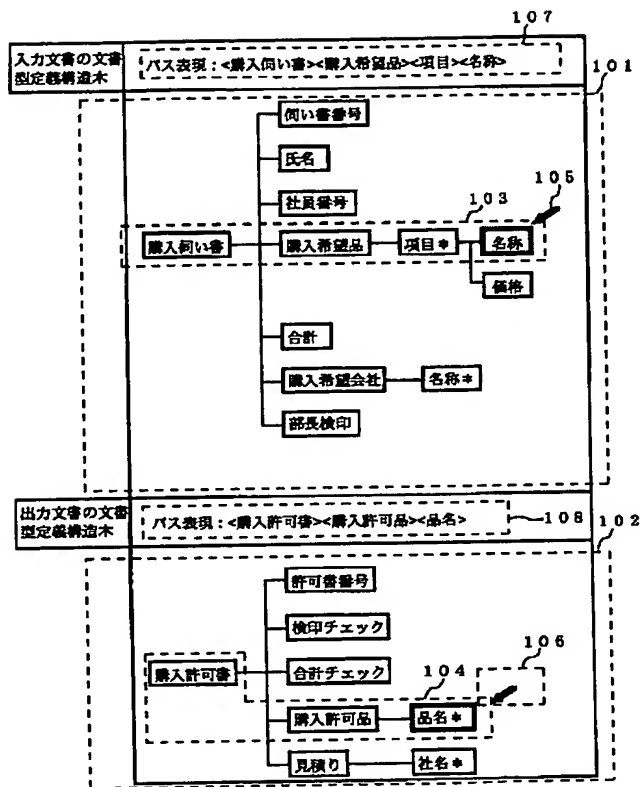
【図54】

チェックID	チェック要素	チェック条件		修正方法
		チェック種別	チェック値	
1	箇条書き、項目	要素数	上限=8	なし
2	索引の読み	文字種	検索文字種=漢字	なし
3	索引	長さ	上限=15	なし
4	段落	文字列	検索文字列=猶	修正文字列=なお
5	段落	文字列	検索文字列=#	なし
6	見出し	長さ	上限=50	なし

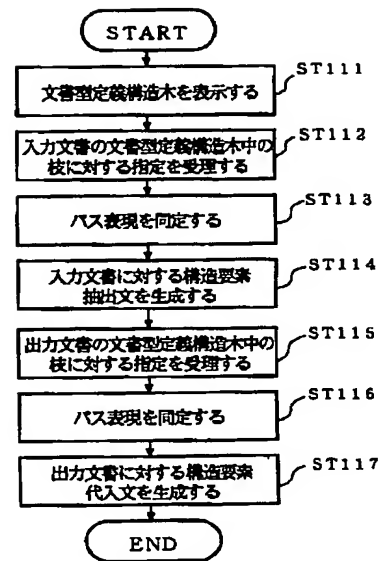
【図47】



【図50】



【図52】



【図53】

